

Vaikka vaimo vaatisi eroa

ero-kantaisen johdospesyeen määrällinen tutkimus

Pro gradu -tutkielma
Joonatan Tola
Helsingin yliopiston
suomen kielen, suomalais-ugrilaisten
ja pohjoismaisten kielten
ja kirjallisuuksien laitos
Lokakuu 2015

Sisällysluettelo

1 JOHDANTO	1
1.1 Aihe synkronisen johto-opin valossa.....	2
1.2 Aineisto, sen rajaus ja esikäsittely	4
1.3 Tutkimuksen tausta	6
1.4 Tutkimuskysymykset ja tutkielman rakenne	8
2 TUTKIMUKSEN TEOREETTINEN TAUSTA	10
2.1 Määrällisen kielentutkimuksen periaatteet	10
2.1.1 Korpuksen rooli työssä	10
2.1.2 Korpusaineiston keruu ja käsittely	11
2.2 Työn merkityskäsitys.....	14
2.2.1 Kategorisointi.....	15
2.2.3 Monimerkityksisyys ja sen kuvaaminen.....	16
2.2.2 Skeemat ja prototyyppisyys	19
2.3 Taustaa muuttujien koodaukselle	22
2.4 Käyttäytymisprofiilit.....	25
3 <i>ERO</i>-PESYEEN MONIMERKITYKSISYYDEN TYÖHYPOTEESIT	29
3.1 Sanakirjojen luokitukset	29
3.1.1 <i>ero</i> -lekseemi sanakirjoissa	30
3.1.2 <i>erota</i> -lekseemi sanakirjoissa	30
3.1.3 <i>erottaa</i> -verbi sanakirjoissa	31
3.1.4 Yhteenveto	32
3.2 Taustatiedon pohjalta luodut skeemat	32
3.2.1 <i>ero</i> -nominin skeemat.....	32
3.2.1 <i>erota</i> -verbin skeemat.....	33
3.2.1 <i>erottaa</i> -verbin skeemat.....	33
3.3 Korpusesimerkkien koodaus skeemoihin	34

3.3.1 Konkreettisen liikkeen ja suhteen katkaisemisen skeemat	35
3.3.2 Laadun ja määrän eroavaisuuden skeemat	36
3.3.3 Abstraktit, kognitiiviset ja aistilliset skeemat	37
3.3.4 <i>ero</i> -nominin mentaaliset skeemat	41
4 MUUTTUJAT	43
4.1 X-, Y- ja Z-argumentit	44
4.1.1 <i>erota</i> - ja <i>erottaa</i> -verbien koodausratkaisut	44
4.1.2 <i>ero</i> -nominin koodausratkaisut	46
4.1.3 Muuttujien X_SEM, Y_SEM ja Z_SEM alakategoriat	53
4.1.4 Muuttujien X_MUOTO, Y_MUOTO ja Z_MUOTO alakategoriat	57
4.2 Argumenttirakennetta koskevat muuttujat.....	62
4.2.1 RAKENNE.....	63
4.2.2 RESIPROOKKI.....	66
4.3 Verbejä koskevat muuttujat	68
4.3.1 VERBI_RAKENNE	69
4.3.2 PÄÄLUOKKA	73
4.3.3 MODUS	73
4.3.4 TEMPUS	74
4.3.5 MODUS1 ja MODUS2	74
4.3.6 KIELTO	77
4.4 Adverbiaaleja koskevat muuttujat	77
4.4.1 SYY	78
4.4.2 KVANTTORI	79
4.4.3 VASTAUS	81
4.4.4 AIKA.....	83
5 YKSIMUUTTUAJAINEN ANALYYSI JA SEN TULOKSET	84
5.1 X^2 -riippumattomuustesti	85
5.2 <i>erota</i> -verbi	86

5.2.1 Muuttujat Y_SEM ja Y_MUOTO	87
5.2.2 Muuttujat Z_SEM ja Z_MUOTO	94
5.2.3 Muuttujat RAKENNE ja RESIPROOKKI	97
5.2.4 Yhteenveto	100
5.3 erottaa-verbi	108
5.3.1 Muuttujat Y_SEM ja Y_MUOTO.....	108
5.3.2 Muuttujat Z_SEM ja Z_MUOTO	111
5.3.3 Muuttujat RAKENNE ja RESIPROOKKI	114
5.3.4 Muuttujat X_SEM ja X_MUOTO.....	119
5.3.5 Skeemojen kuvaus analyysin perusteella	122
5.3.6 Yhteenveto ja monimuuttujainen näkökulma	132
5.4 ero-nomini	134
5.4.1 Muuttujat Y_MUOTO ja Y_SEM	135
5.4.2 Muuttujat Z_MUOTO ja Z_SEM	137
5.4.3 Muuttujat X_SEM ja X_MUOTO	139
5.4.4 Muuttujat RAKENNE ja RESIPROOKKI	140
5.4.5 Yhteenveto	143
5.5 Yksimuuttujaisen analyysin koonti	151
6 MONIMUUTTUJAINEN ANALYYSI JA SEN TULOKSET	156
6.1 Agglomeratiivinen hierakkinen klusterianalyysi.....	156
6.2 ero-pesyeen klusterianalyysi	160
6.2.1 Klusterianalyysi A.....	161
6.2.2 Klusterianalyysi B	168
6.2.3 Yhteenveto klusterianalyysistä.....	171
7 PESYESKEEMOJEN MÄÄRITTÄMINEN	174
8 LOPUKSI	180
LÄHTEET.....	189
LIITE 1: HERKKYYSTAULUKOITA	198

1 Johdanto

Suomen kielelle ominaista on erilaisten sananmuodostuskeinojen runsaus ja etenkin johtaminen – ”kielen suurin rikkaus ja ihanin ominaisuus” (Lna). Johto-oppia, tai sen morfologista puolta, onkin tutkittu fennistiikassa paljon. Useimmat kieliopit Setälästä (1880) lähtien ovat esittäneet muodolliset kriteerit johdosten päätyypeille, tuoreimpina Iso suomen kieliooppi (2004, jatkossa VISK). Varsinaisia semanttisia johto-opin tutkimuksia on kuitenkin vähän (ks. Jääskeläinen 2004), eikä tiedossani ole, että johdospesyeen semantiikkaa olisi ainakaan työni mittaluokassa tutkittu määrällisin menetelmin ja korpusaineistoa hyödyntäen.

Pro gradu -työssäni tarkastelen *ero*-kantaisten johdospesyeen lekseemien *erottaa*, *erota* ja *ero* monimerkityksisyyttä HS.fi-korpusaineiston ja määrällisten analyysimenetelmien avulla. Vaikka työni on määrällinen, työ sisältää myös laadullisia vaiheita, joita ovat esimerkiksi muuttujien luominen, aineiston merkitseminen sekä tulosten tulkinta. Laadullista kielentutkimusta teen erityisesti esitellessäni koodausratkaisuja. Merkitysentutkimus lieneekin mahdotonta ilman tutkijan kieli-intuition osallisuutta.

Tarkastelen johtamista synkronisena ilmiönä eli tutkin kieltä tästä hetkestä käsin. Koska työni on synkroninen, en ota kantaa diakronisiin kysymyksiin, esimerkiksi derivaa-tion suuntaisuuteen, siihen kumpi oli ensin, kanta vai johdos, *ero* vai *erottaa*. En tarkastele kannoista johtimilla johdettujen johdosten morfosemanttisia suhteita tai johdostyyppien muodollisia kriteereitä. Tarkastelen johdoksia keskenään tasavertaisina leksikaalisina elementteinä, joilla tuotetaan kielellisiä merkityksiä ihmisten välisessä vuorovaikutuksessa. Analyysissä keskityn yksinomaan kielen semanttiseen puoleen, *ero*-lekseemien välisiin semanttisiin kytköksiin.

Työtäni voisi luonnehtia kokeelliseksi. *ero*-kantaisen johdospesyeen monimerkityksisyyden kuvauksen lisäksi tutkimukseni tavoitteina on selvittää, soveltuvatko määrälliset analyysimenetelmät ylipäänsä johdospesyeen ja nimenomaan *ero*-pesyeen semanttiseen analyysiin. Työ sisältää kaksi erillistä määrällistä analyysivaihetta. Yksimuuttujaisessa vaiheessa testataan aineistoa nollahypoteesitesteillä. Monimuuttujaisessa vaiheessa aineistoa analysoidaan hierarkkisen klusterianalyysin avulla. Monimuuttujainen analyysi pohjaa niin sanottuihin käyttäytymisprofileihin (Gries 2006, 2012). Ne heijastavat kunkin skeeman eli ”merkityksen” semantis-syntaktista profiilia, joka kumpuaa aktuaalisesta kielenkäytöstä. Fennistiikassa ei tietääkseni ole tutkittu derivaatiota käyttäytymisprofileihin pohjautuvalla

menetelmällä. Määrälliset analyysit on toteutettu *R*-ohjelmointikielen (jatkossa *R*-kieli, -ohjelma tai *R*) skripteillä, jotka ovat Arppen (2015), Griessin (2010) ja itseni kirjoittamia.

Työni teoreettisena kivijalkana seisoo näkemys siitä, että kielellisen ilmaisun merkitys voidaan selvittää kontekstuaalisten yhdessä esiintymien nojalla tarkasteltavan ilmiön jakaumasta, distribuutiosta (ks. Divjak & Gries 2006: 24–28). Tämä merkityskäsitys nousee Firthin (1957) muodostamasta traditiosta, jossa merkitys ymmärretään kontekstuaalisesti vastakohtana (de)kompositionaalisuudelle.

Työn teoreettisena paradigmana toimii kognitiivinen kielentutkimus. En kuitenkaan kannan teorian lippua orjallisesti, vaan teen tutkimusta pikemminkin kognitiivisessa hengessä. Tutkimus edustaa löyhästi kognitiivista korpuslingvistiikkaa, ja on jatkoa laadulliselle kandidaatin tutkielmalleni, jossa tutkin *ero*-pesyeen semantiikkaa 1990-luvun uutiskielessä.

1.1 Aihe synkronisen johto-opin valossa

Johtamista voidaan tutkia joko diakronisena tai synkronisena ilmiönä (Lehtinen 1979: 244), mutta tyypillisesti johtaminen käsitetään nimenomaan historialliseksi prosessiksi (ks. Koivisto 1991: 9)¹. Tämä on ymmärrettävää, sillä jo johto-opin käsitteeseen sisältyy suuntaisuuden oletus ja siten historiallinen implikaatio: johtaminen on temporaalinen prosessi, jossa tunnetusta luodaan uutta (Jääskeläinen 2004: 53).

Tässä tutkimuksessa tarkastelen johtamista kuitenkin pelkästään synkronisena ilmiönä, jolloin johtosuhteissa keskeistä on se, miten sanojen nähdään motivoivan toisia sanoja nykykielessä (ks. Kytömäki 1977: 28–32). Synkronisesti johtamisessa on toisin sanoen kyse lekseemien välisestä korrelaationsuhteesta, jota ei ajatella suuntaiseksi (Jääskeläinen 2004: 53). Tällöin kantasana (*ero*) ja johdokset (*erottaa* ja *erota*) voidaan nähdä toistensa korrelaateiksi (ks. Lehtinen 1979: 245). Korrelaation sijaan käytän työssäni assosiaation käsitettä, jolla korostan sitä, että *ero*-lekseemien (tai -semeemien) välinen semanttinen suhde ei ole (synkronisessa tarkastelussa) kausaalinen tai temporaalinen vaan assosiatiivinen.

Vaikka tarkastelutapani on synkroninen, johdokset lekseemeinä tai ainakin morfologisina entiteetteinä ovat kielen historiallisen kehityksen tuotoksia. Morfosemanttisen segmentaation näkökulmasta *ero*-kantisilla johdoksilla *erottaa* ja *erota* on sama etymologinen kantasana, nimittäin *ero*-nomini, jonka nojalla *ero*-lekseemit kytkeytyvät samaan johdospesyeen. Käsitteellä kantasana tarkoitan lekseemiä, josta johdos on johtimen avulla historiallisesti ajateltuna muodostettu ja jonka vartalo on osa johdoksen vartaloa (ks. VISK §

¹ Synkronisena prosessina voidaan Räisäsen (1988: 16) mukaan tarkastella vain produktiivista johtoa.

157). Johtimella tarkoitan suffiksia eli itsenäisenä sanana esiintymätöntä kielenainesta, joka kiinnittyy sanavartaloon ja muodostaa sen kanssa uuden lekseemin eli johdoksen (VISK § 156). Koska *erottaa-* ja *erota-*verbien kantasana on *ero-*nomini², ne ovat demoninaalisia eli nominikantaisia johdoksia: transitiiiverbi *erottaa* on johdettu *ero-*kantanominista *ttA-*kausatiivijohtimella (*ero+ttA* > *erottaa*), intransitiivinen supistumaverbi *erota* *tA-*johtimella (*ero+tA* > *erota*) (VISK § 305)³.

En tutkimuksessani puutu *ero-*kantaisen johdospesyeen⁴ morfologiseen jäsentymiseen tämän enempää. Koska tutkimukseni on synkroninen ja koska keskityn johtamisessa pelkästään semantiikkaan, tarkastelen kantasanoja ja johdoksia jatkossa täysin samanarvoisina kielellisinä yksikköinä ottamatta kantaa johtamisen suuntaisuuteen. Yksisuuntaista $a > b > c$ -johtamista eli ns. johtimella johtamista, jossa kielitajussa olevaan irralliseen kantaan liitetään kielitajussa oleva irrallinen johtosuffiksi, voidaankin kritisoida. Esimerkiksi Räisänen pitää sitä johtamisen muotona, jota ”harrastavat ainakin lingvistit” (Räisänen 1978: 338)⁵. Jääskeläinen (2004: 54) taas toteaa, että milloin pesyeen jäsenet ketjuuntuvat morfosemanttisesti ideaalin rypään⁶ mukaan, on merkityksetöntä, missä järjestyksessä tai millä tavoin pesyeen jäsenet ovat historiallisesti toisiinsa kytkeytyneet.

Viittaan jatkossa käsitteillä *ero-*nomini, *erota-*verbi ja *erottaa-*verbi, lekseemeihin *ero*, *erota* ja *erottaa* ilman minkäänlaista historiallis-temporaalista implikaatiota. Käsitteellä *ero-*skeema viittaan lekseemeittäin erittelemiini merkitysvariantteihin. Käsitteillä *ero-*johdokset tai *ero-*lekseemit viittaan joukkona *ero-*lekseemipopulaatioon eli *ero-*, *erota-* ja *erottaa-*lekseemeihin. Vaikka tutkin johdospesyettä, en tutki ”johdoksia” vaan lekseemejä tai pikemminkin semeemejä – etymologisesti *ero-*kantaisten sanojen ”eri merkityksiä” ja näiden eri merkitysten välisiä semanttisia ja morfosyntaktisia yhtäläisyyksiä ja eroja, jotka ovat hypoteesini mukaan paikannettavissa kontekstista. Tutkimukseni keskiössä eivät (paradoksaalisesti) olekaan yksittäiset *ero-*sanat, vaan sanat, muodot ja rakenteet *ero-*sanojen ympärillä.

² Vallalla olevan etymologian mukaan *ero-*nomini on johdettu *erä-*pesyeessä (ks. Häkkinen 2007).

³ VISK:n (§ 305) mukaan *erottaa-*verbi voidaan tulkita joko nomini- tai verbikantaiseksi (<*erota* tai <*ero*). Tulkinta on ilmeisesti väärä (sivuan asiaa tässä työssäkin). Sekä *erottaa* että *ero* (myös kirjoitusasussa *eroittaa* ja *eroitus*) ovat esiintyneet suomen kielessä Agricolasta lähtien. *erota-*lekseemi sen sijaan on tietoisesti 1800-luvulla luotu uudissana, joka on korvannut *erkaantua-* ja *eritä-*verbien merkitystehtäviä. (Häkkinen 2007; SSA.) Kielenkäytön kannalta deverbaalisuudella tai denominaalisuudella ei tosin ole merkitystä, vaan sanojen välisellä korrelaatio-suhteella (Ojutkangas 1997: 361).

⁴ Johdospesyeen rinnalla kirjallisuudessa käytetään käsitteitä *sanaryhmä*, *sanaperhe*, *sanepesye* ja *sanue*.

⁵ Johtimella johtaminen on vain yksi johtotapa erilaisten analogiaan perustuvien johtotapojen rinnalla (korrelaatio-, mallisana- ja takaperoisjohto) (ks. Räisänen 1978; implisiittisesti Penttilä 1957: 273; kootusti Laakso 1990: 9–18).

⁶ On epäselvää, mitä Jääskeläinen tarkoittaa ideaalilla rypäällä. Onko *ero-*pesye ideaali rypäs? En osaa sanoa.

1.2 Aineisto, sen rajaus ja esikäsittely

Pitäkseni aineiston kohtuullisen kokoisena rajasin tutkimukseni ulkopuolelle ison joukon *ero*-kantaisia sanoja, esimerkiksi lekseemit *erottua*, *erotella*, *erottuva* ja *erottumaton* sekä lekseemit, jotka eivät ole etymologisesti *ero*-kantaisia, kuten *erotiikka*. Työni kartoittaa *ero*-kantaisen johdospesyeen polysemiaa vain tarkastelemieni kolmen lekseemin (*ero*, *erota*, *erottaa*) osalta ja rajoittuu valitsemaani korpusaineistoon eli tiettyyn rekisteriin. Tutkimustulokseni pätevät vain valitsemaani aineistoon, eikä niitä tietenkään voi suoraan yleistää muihin (vrt. Karlsson, Hakulinen & Vilkuna 1980: 94).

Kognitiivisen kielentutkimus lähtee siitä, että tarkastellaan käyttöpohjaista kieltä, mikä käytännössä usein tarkoittaa korpusaineiston hyödyntämistä (ks. Onikki-Rantajääskö 2001). Aineiston käyttö kielenanalyysissä ei ole itsestäänselvyys. Esimerkiksi Talmy (2005) on puolustanut kielen tarkasteluun kouluttautuneen kielentutkijan oman kielitajun ja introspektion käyttöä kielentutkimuksessa. Tummersin, Heylenin ja Geeraertsin (2005: 232) mukaan korpus voittaa introspektion siinä, että se tuottaa aktuaalisesta kielenkäyttötilanteesta taltioitua spontaania dataa. Ojutkangas (2008: 277) perustelee empiirisen aineiston käyttöä mielikuvituksen puutteella: tutkija ei pysty mielikuvituksen ja kieli-intuitionsa avulla kuvaamaan kieltä riittävän tarkasti. Hän kuitenkin toteaa, etteivät aineistoon pohjautuvat ja introspektiopohjaiset lähestymistavat sulje toisiaan pois vaan pikemminkin täydentävät toisiaan. Empiirisen tutkimusotteen ja niin sanotun nojatuoli-lingvistiikan vastakkaisasettelu voidaan kin kyseenalaistaa, sillä empiirinenkin tarkastelu ei ole vapaa tutkimuksen subjektiivisuudesta, joka voi näkyä esimerkiksi aineiston valinnassa, sen analysoinnissa ja tulosten arvioinnissa (Huomo & Sivonen 2010: 38; Itkonen & Pajunen 2010: 58; Salminen 2012: 16).

Aineistoni on peräisin Kielipankin Korp-palvelimelta (corp.csc.fi) HS.fi-uutiskommenttiaineistosta (jatkossa HS.fi-korpus), jonka tekstinäytteet ovat vuodelta 2011–2012. HS.fi-korpus on rekisteriltään siinä mielessä sekakoosteinen, että se sisältää paitsi uutisartikkelit myös uutisia kommentoineiden ihmisten keskustelujaksot. HS.fi-korpus valikoitui aineistoksi paitsi korpuksen tuoreuden vuoksi myös sen takia, että korpus edustaa suhteellisen spontaanisti, aidossa kielenkäyttötilanteessa synnytettyä kirjoitettua (puhekielistä) kieltä. Toimituksen tuottamien uutisartikkeleiden osuus tekstinäytteistä on 7.5 %:a (*erottaa*-verbin tapauksessa). Loppu on ”tavallisten talliaisten” tuottamaa.

Rakensin aineistoni poimimalla Korp.fi-käyttöliittymän avulla HS.fi-korpuksesta kaikki lauseet, joissa *erottaa*- tai *erota*-verbi esiintyy joko persoonamuodossa (*erottaa*

n=260⁷, *erota* n=399), tai A-infinitiivin perusmuodossa (*erottaa* n=210, *erota* n=164) sekä lauseet, joissa *erottaa*-verbi esiintyy MA-infinitiivin illatiivissa (*erottaa* n=39). *erottaa*- ja *erota*-verbien MA-infinitiivin muilla sijamuodoilla on HS.fi-korpuksesta vain muutamia esiintymiä. *erota*-verbin MA-infinitiivin illatiivilla on HS.fi-korpuksesta 35 esiintymää, jotka rajasin tutkimuksestani tasapainottaakseni *erottaa* ja *erota* verbien esiintymien kokonaislukumäärää (*erottaa* n=509, *erota* n=563). Havaintojeni mukaan mainitsemiäni MA-infinitiivimuotojen poisjättäminen ei vaikuta tutkimukseni tuloksiin.

ero-verbien lisäksi poimin HS.fi-korpuksesta kaikki lauseet, joissa *ero*-nomini (n=594) esiintyy joko yksikön nominatiivissa (n=365) tai yksikön partitiivissa (n=229). Aikomukseni oli sisällyttää aineistooni myös *ero*-nominin yksikön ja monikon genetiivimuodot⁸ sekä nominatiivin ja partitiivin monikkomuodot, mutta rajasin ne ulos yksinkertaistaakseni aineistoani. Samasta syystä jouduin rajaamaan pois laajafrekvenssiset inessiivi- ja illatiivimuodot jotka esiintyvät rakenteissa *päästä eroon* ja *olla erossa*. Rakenteet, joissa *ero*-nomini esiintyy yhdyssanan osana, olen rajannut pois.

Partitiivin valitsin, koska siinä (ja genetiivissä jonka rajasin ulos) *ero*-nomini esiintyy sellaisten havainto- ja kognitioverbien kuin *maistaa* ja *tietää* objektina. Nämä assosioituvat eräisiin *erottaa*-verbin käyttötapoihin. Ilmiötä havainnollistavat aineistoesimerkit 1a ja 1b:

1a. **Vaikeahan** sitä **eroa on maistaa** kun ovat vedellä lantrattuja tarkkelysliemiä. (K1100)

1b. Peruslageroluiden **makua ei voi erottaa**⁹ sokkoteistissä. (K143)

Esimerkissä 1a *ero*-nomini esiintyy *maistaa*-verbin ja 1b:ssä *maku*-nomini *erottaa*-verbin objektina. Yhteistä näille on, että molemmat ilmoittavat jonkinlaista aistikokemusta.

erottaa-verbin rajasin tutkimuksestani pois verbin suhteellisen niukkafrekvenssisyyden vuoksi (n=96). Se oli harmi, sillä havaintojeni mukaan sekä *erottaa*-verbi että *ero*-nomini esiintyvät eräissä *erottaa*-verbin merkitystehtävissä. Arppen (2008: 65) hengessä en poiminut *erota*- ja *erottaa*-verbien populaatiosta satunnaisotosta, esimerkiksi n=100 kuten

⁷ Kirjain n, esimerkiksi n=10, tarkoittaa, että havaintoja on 10 kappaletta. Työssä poiketaan nykysuomen oikeinkirjoitusnormeista (ks. KTOO: 75) siten, että yhtäsuuruusmerkit – joita työssä on 1094 kappaletta – merkitään ilman välilyöntejä, kun kyse on *df*- tai *a*-arvosta (*df*=15 ja $\alpha=0.05$) tai assosiaation voimakkuuden mittaustuloksesta ($U_{\text{muuttujajakskeema}}=0.389$). Skeemoihin ja luokkiin kytkeytyvät arvot merkitään kuitenkin normien mukaisesti ($Z_{\text{MUOTO_LÄSNÄ}} = 60.7\%$). Lisäksi käytän desimaalierottimena pilkun sijaan pistettä.

⁸ Genetiivimuodot olisivat olleet kiinnostavia siksi, että yksi *ero*-nominin yleisimmistä lauseenjäsennyspaikoista on esiintyä sellaisten transitivisten kognitio- ja aistimisverbien kuin *ymmärtää* ja *huomata* objektina. Tällöin totaali- ja partitiiviobjektin oppositio ilmenee muotojen *eroa* ~ *eroja* ja *eron* ~ *erojen* vaihteluna.

⁹ Olen lihavoinut korpusesimerkeistä *ero*-, *erota*- tai *erottaa*-lekseemin sisältävät verbilausekkeet. Alleviivausta käytän korostaakseni tarkkailtavaa kielenpiirrettä. Suuraakkosia käytän skeemojen (EROTA.IRTI, ERO.ERI) ja muuttujien nimissä (RESIPROOKKI, NA). Korpusesimerkeissä muuttujien luokkia esitän supertekstinä, kuten *Mä*_{X-ARG}. **erotin** *papanoita*_{Y-ARG}. *lattialla*.

Biber, Conrad ja Susan (1998: 92) neuvovat. Sen sijaan otin rajauksieni puitteissa kaikki esiintymät siitäkin huolimatta, ettei HS.fi-korpukseen ole koodattu morfosyntaktisia saati semanttisia tunnisteita. *ero*-nominin analysointia aloittaessani *erottaa*- ja *erota*-verbien suurimmaksi osaksi käsin tekemäni annotointi oli viisastuttanut minua sen verran, että sovelsin poimintaan jonkinmoista satunnaisotantaa. Umpimähkäisesti valittu otoskoko (n=594)¹⁰ sisältää *ero*-nominin sijamuotoja siinä suhteessa kuin niitä on HS.fi-korpuksessa (jossa yksikön nominatiivin n=577 ja yksikön partitiivin n=361 eli yhteensä n=938)¹¹. Merkitsin kullekin tekstiesimerkeille satunnaisluvun, jonka avulla arvoin sijoja suhteessa esiintymistajuuteen niin, että partitiivia on 38.5 % (n=229) ja nominatiivia 61.5 % (n=365).

Lopullisessa aineistossani on täten 1666 *ero*-lekseemiä seuraavassa suhteessa: *ero* 35.6 % (n=594), *erota* 33.8 % (*erota* n=563) ja *erottaa* 30.6 % (*erottaa* n=509). HS.fi-korpuksesta poiminani lauseet – tai tietueiden sisältämät *ero*-lekseemit koteksteineen – toimivat tutkimukseni tilastoyksikköinä¹². Tutkimukseni tekstiesimerkit ovat peräisin aineistostani, jolla viittaan tähän 1666 esimerkin kokoelmaan, tutkimuskorpukseeni. En anonymiteetin vuoksi käytä työni korpuksimerkeissä HS.fi-uutiskommenteissa olevia nimimerkkejä tai HS.fi-korpuksen tietueiden ID-tunnisteita, vaan olen arponut kullekin tekstinäytteelle oman tunnusteen. Tekstinäytteet, jotka saavat arvot K1–K509, sisältävät *erottaa*-verbin, K510–K1062 *erota*-verbin ja K1073–K1666 *ero*-nominin. Mikäli en ilmoita tekstinäytteen kohdalla K-tunnusta, olen (yleensä) keksinyt esimerkin havainnollistaakseni jotain ilmiötä.

Tutkimuskorpuksen keräämisen jälkeen syötin aineiston taulukointiohjelmaan (*Open Office Calc*), jonne loin merkitsemissysteemin, jonka muuttujista tässä työssä on mukana 19. Tämän jälkeen merkitsin eli annotoin kullekin tilastoyksikölle arvon suhteessa kuhunkin muuttujaan. Yhteensä koodasin 31654 muuttujan arvoa. Muuttujiston luomisprosessin selitän seikkakohtaisesti luvussa 4. Selvitän työni kulkua yksityiskohtaisesti vielä luvussa 1.4, mutta sitä ennen tarkastelen pro gradu -tutkimukseni taustoja kandidaatin työni valossa.

1.3 Tutkimuksen tausta

Tutkin *ero*-pesyettä laadullisessa kandidaatin tutkielmassani (Tola 2010), joka käsitteli *erottaa*- ja *erota*-verbien (ja kursiivisesti myös *ero*-nominin) polysemiaa tekstinäytteissä, jotka

¹⁰ Otoksoon suuruuteen vaikutti *erota*- ja *erottaa*-verbin otoskoot. Pyrin samankokoisiin otoksiin.

¹¹ HS.fi-korpuksessa yksikön ja monikon partitiivin, nominatiivin ja genetiivin yhteenlaskettu n=1438, josta yksikön genetiivin (n=129), monikon genetiivin (n=19), yksikön partitiivin (n=361), monikon partitiivin (n=204), yksikön nominatiivin (n=577) ja monikon nominatiivin (n=148).

¹² Tilastoyksikkö on se tilastollinen perusyksikkö, jota tilastointi koskee tai jolta tilastoitavat tiedot kerätään, esim. Suomen kansalainen, kotitalous tai kunta (Tilastokeskus 2015).

oli poimittu lähinnä sanomalehti Karjalaisen vuosikerroista 1998–1999. Kandidaatin tutkielman löydökset ovat ainakin implisiittisesti ohjanneet pro gradu -työn suunnittelua ja toteutusta. *ero*-pesyeen parissa työskentelyllä on ollut vaikutusta ainakin käsityksiini siitä, kuinka pesyeen polysemia rakentuu, mikä heijastunee skeemojen erittelyssä (luvussa 3).

Kandidaatin tutkielmassani määritin eri ”semeemi-” tai merkitystyyppijä *erottaa*-verbille 13, *erota* verbille 7 ja *ero* nominille 10. Merkitystyyppit kuvaavat lekseemien erilaisia käyttötapoja, kuten *erota kirkosta* tai *erota parisuhteesta*. Keskeisin havaintoni oli, että eri merkitystyyppit eivät ole toisistaan erillisiä, vaan päinvastoin merkitysvarianteista punoutuu semanttisia jatkumoa, jotka ylittävät lekseemien väliset ”morfologiset” rajat. Tätä merkitysketjujen jatkumoa tai semeemien kytköksisyyttä kuvasin merkitysverkolla, jonka idea on visualisoida se, että *ero*-sanoilla ilmaistaan samankaltaisia kielenulkoisia tilanteita.

Kandidaatin tutkimuksessani esitän neljä päämerkitysketjua. Ne kuvaavat parisuhteen katkeamista (A-ketju), työsuhteen päättymistä (B-ketju), ominaisuuden tai laadun poikkeavuutta (C-ketju) sekä aistillista toimintaa (D-ketju). Osa ketjuista on ”täydellisiä”, jolloin niillä on edustus jokaisessa *ero*-lekseemissä (B- ja C-ketju), osa vajaita, jolloin vain osa *ero*-lekseemeistä koodaa ketjun ilmaisemaa merkitystä (A- ja D-ketju).

A-ketju, parisuhteen katkeaminen: *Matille ja sen muijalle tuli ero; Matti erosi muijastaan.*

B-ketju, työsuhteen päättyminen: *Matin ero työstä; Matti eroaa työstä; Keijo erottaa Matin työstä.*

C-ketju, ominaisuuden poikkeavuus: *Kissan ja koiran ero on, että kissa on itsenäinen; Kissa ja koira eroaa siinä, että kissa on itsenäinen; Kissaa ja koira erottaa se, että kissa on itsenäinen.*

D-ketju, aistillinen toiminta: *Huomata ero kissan ja koiran välillä; Erottaa kissa ja koira toisistaan.*

Näiden lisäksi erittelin joukon merkityksiä, jotka esiintyvät lekseemikohtaisesti.

Havaitsin, ettei *erota*-verbi lainkaan koodaa mentaalista toimintaa. Sen sijaan *erota*-verbi kuvaa agenttiivisia prosesseja, joissa ihminen joko fyysisesti tai ”metaforisesti”¹³ liikkuu positiosta toiseen. *erottaa*-verbi taas kuvaa prosesseja, joissa ihminen energettiseen aloitteeseen kykenevän olennon vaikutuksesta kokee muutoksen (B-ketju) sekä tiloja, joissa kokija havainnoi päänsisäistä tai ulkoista todellisuutta (D-ketju)¹⁴. Havaitsin lisäksi, että *eronomini* yhtyy useisiin verbeihin – kuten *ottaa ero*, *tehdä ero*, *päästä eroon*, *huomata ero* – päästen näin predikoimaan sekä *erota*- että *erottaa*-verbin merkitystehtäviä.

¹³ Työ nojasi lokalistiseen hypoteesiin, jonka keskiössä on spatiaalinen liike ja sijainti ja joka johtaa muunlaiset semanttiset suhteet – sosiaaliset, abstraktit jne. – spatiaalisuudesta (ks. Onikki-Rantajääskö 2001: 224–226).

¹⁴ Patientti voi olla myös eloton olio.

Sivuutan tässä pro gradu -työssä kandidaatin tutkielmani merkitysjaottelun ja otan *ero*-lekseemien semeemijaottelun lähtökohdaksi sanakirjojen, subjektiivisen aineistoanalyysin ja intuitioni pohjalta tekemäni uuden jaottelun, jonka esittelen luvussa 3.

1.4 Tutkimuskysymykset ja tutkielman rakenne

Tutkimukseni tavoite on määrällisin eli kvantitatiivisten analyysimenetelmien avulla selvittää, vastaako korpusaineistossa havaittu kieli sanakirjoihin ja intuitioni pohjautuvaa *erottaa*-, *erota*- ja *ero*-lekseemien hypoteettista skeemajaottelua. Tavoitteena on siis:

- i) selvittää, tukeeko määrällinen analyysi oletusta, että kunkin *ero*-lekseemin semantiikassa esiintyy polysemiaa;
- ii) kuvata, miten *ero*-pesyeen polysemia rakentuu *erottaa*-, *erota*- ja *ero*-lekseemien merkitysavaruuksissa valitsemieni rajoitusten puitteissa. Yksittäisten lekseemien kuvauksen lisäksi tavoitteeni on selvittää, onko *ero*-lekseemien skeemojen välillä semanttisia kytköksiä myös johdospesyeen tasolla;
- iii) tavoitteeni on lisäksi selvittää, mistä *ero*-skeemojen semanttiset erot johtuvat, toisin sanoen, paikantaa muuttujat, jotka kytkeytyvät skeemojen merkitysvaihteluun lekseemien sisällä (yksimuuttujainen analyysi) ja lekseemien välillä (monimuuttujainen analyysi);
- iv) lisäksi tutkimukseni läpi kulkevana (tausta)tavoitteena on selvittää, voiko johdospesyeen polysemiaa ylipäänsä tutkia valitsemillani määrällisillä menetelmillä ja toisaalta, onko polysemiaa sellaisena ilmiönä kuin sen määrittelyn edes olemassa tutkimassani *ero*-pesyeessä.

Työni rakentuu viidestä kokonaisuudesta: teoriasta (luku 2), skeemojen esittelystä (luku 3), muuttujien esittelystä (luku 4), aineistoanalyysistä (luvut 5 ja 6), pesyeskeemojen määrittelystä eli tutkimustulosten raportoinnista (luku 7) ja loppupäätelmistä (luku 8). Aluksi luvussa 2 luon silmäyksen työni kannalta oleelliseen teoriaan. Tutustun määrälliseen tutkimusotteeseen, määrällisessä tutkimuksessa käytettävien muuttujien periaatteisiin, käytäytymisprofileihin, kognitiivisen kielentutkimuksen merkityskäsitykseen sekä monimerkityksesisyyteen ja sen kuvaamiseen.

Luvussa 3 selostan, miten sanakirjat jaottelevat *ero*-lekseemien merkitysrakennetta, minkä jälkeen selostan, miten itse intuition ja sanakirjojen pohjalta jaottelen *ero*-lekseemien merkitysrakennetta. Lopuksi esitän kriteerit, joiden perusteella annotoin kullekin aineistoesimerkille tiedon, mihin skeemaan se lekseemin sisäisessä merkitysavaruudessa lukeutuu.

Luvussa 4 esittelen tutkimuksessani käytettävät muuttujat ja niiden koodausprosessin.

Muuttujia on kolmenlaisia: semanttisia, morfologisia ja syntaktisia. Työssäni tarkastelen etenkin *ero*-lekseemieni täydennysten rakenteellisia ja semanttisia piirteitä. On syytä pitää mielessä, että koko määrällisen analyysin laatu riippuu ratkaisevasti analyysiä edeltävästä koodaustyöstä (ks. Pitkänen & Kohonen 1984: 92; Karlsson yms. 1980: 6–8), joka yksinkertaistaen käsittää kolme vaihetta: 1) muuttujien valinnan, 2) muuttujien alakategorioiden valinnan ja 3) muuttujien arvojen merkitsemisen eli annotaatioprosessin.

Työni sisältää kaksi analyysivaihetta, yksimuuttujaisen vaiheen (luku 5) ja monimuuttujaisen vaiheen (luku 6). Tutkimukseni pääpaino on yksimuuttujaisessa analyysissä, joka rakentuu kolmesta ”erillistutkimuksesta”: *erota*-verbin (luku 5.1), *erottaa*-verbin (luku 5.2) ja *ero*-nominin (luku 5.3) analyysistä. Näissä analyyseissä testaan yksittäisten muuttujien ja skeemojen välisiä assosiaatiosuhteita khiin neliö -testillä sekä niin sanotuilla *post hoc* -testeillä. Vaiheen tarkoitus on selvittää, mitkä muuttujien alakategorioista kytkeytyvät mihinkin skeemaan, toisin sanoen, mitkä alakategorioista parhaiten selittävät *ero*-skeemojen kontekstuaalista käyttäytymistä (vrt. Arppe 2008: 73). Analyysissä keskityn X-, Y- ja Z-argumenteiksi nimittämiini *ero*-lekseemien täydennyksiin tai seuralaisiin (ks. luku 2.3).

Monimuuttujaisessa analyysivaiheessa (luku 6) määrällisenä tutkimusvälineenä toimii hierarkkinen klusterianalyysi. Vaiheen tarkoitus on syventää yksimuuttujaisen analyysin tuloksia ulottamalla semanttinen analyysi yksittäisistä lekseemeistä johdospesyeen tasolle. Klusterianalyysissä tutkin, mitkä skeemat ryhmittyvät yhteen ja mitkä erikseen eli mitkä skeemoista edustavat pikemminkin samankaltaista merkitystä ja mitkä erilaista. Monimuuttujainen vaihe on tärkeä, sillä kuten Gries (2003: 79, 107–108) ja Arppe (2008: 256) muistuttavat, kielenkäyttäjät ei tuota aktuaalisessa kielenkäyttötilanteessa kielenaineiksia yksittäisten tekijöiden, vaan lukuisten yhtä aikaa yhdessä toimivien ja toisiinsa kytkeytyneiden tekijöiden pohjalta¹⁵. Monimuuttujainen vaiheen tarkoitus on toimia työni eräänlaisena langat yhteen punovana synteeseinä. Koska monimuuttujainen vaihe työni laajuuden vuoksi kuitenkin on suppea, ensisijaisesti se toimii näkökulmia avaavana kokeiluna, josta toivottavasti nousee horisontteja *ero*-pesyeen – tai jonkin toisen kielenilmiön – jatkotutkimusmahdollisuuksiin.

Esitän tutkimustuloksia paitsi sanallisesti myös taulukoiden, diagrammikuvioiden (luvut 4 ja 5) ja dendogrammikuvien (luku 6) avulla. Luvussa 7 ja 8 luon yhteenvedon tästä kaikesta. Aloitan teorian esittelemisestä.

¹⁵ Gries (2015a: 2) kirjoittaa: “No linguistic phenomenon is ever truly monofactorial in nature, [...] correspondingly, the field has been moving in the direction of multifactorial testing.”

2 Tutkimuksen teoreettinen tausta

Tässä luvussa esittelen työni kannalta oleellista teoriaa. Aloitan tarkastelemalla määrällistä tutkimusotetta, korpuksen roolia työssäni, aineistonkeruumenetelmiä sekä määrällisen työn kriteereitä (luku 2.1). Tutkimukseni merkityskäsityksen ja merkityksen kuvauksen teoreettista taustaa esittelen luvussa 2.2. Luvussa 2.3 taustoitan työssäni käytettyjä muuttujia, minkä jälkeen luon katsauksen käyttäytymisprofiileihin (luku 2.4).

2.1 Määrällisen kielentutkimuksen periaatteet

Tutkimusotteeni on määrällinen ja edustaa niin sanottua kognitiivista korpuslingvistiikkaa, joka on Griessin (2015b: 113) mukaan viime vuosina noussut aivan kielitieteen keskiöön (korpuslingvistiikasta ks. Biber yms. 1998: 1–11; Grondelaers, Geeraerts ja Speelman 2007).

Työn määrällisen luonteen vuoksi myös analyysityökalut ovat määrällisiä: ne perustuvat lukumääriin ja numeerisiin mittatuloksiin, erotuksena laadullisista työkaluista ja menetelmistä, jotka ovat luonteeltaan sanallisia ja painottavat tutkittavan ilmiön niitä puolia, joita ei voi numeerisesti mitata (Luoja 2006: 10). Työkaluja erittelen tarkemmin sitä mukaan kuin sovellan niitä eli luvuissa 5 (yksimuuttujaiset menetelmät) ja 6 (monimuuttujaiset menetelmät). Käyttäytymisprofiilien teoreettista taustaa esittelen kuitenkin jo luvussa 2.4.

Määrälliselle tutkimukselle on ominaista, että konkreettiset havainnot, analyysi, mittaus tulokset ja näiden pohjalta tehdyt päätelmät perustuvat empiiriseen aineistoon, joka on muun kuin tutkijan itsensä tuottamaa (Pitkänen & Kohonen 1984: 3; Nummenmaa 2004: 19). Korpuslingvistiikassa korpuksen käyttö on tietenkin itsestään selvää (ks. Biber yms. 1998).

2.1.1 Korpuksen rooli työssä

Korpuksen rooli eri tutkimuksissa vaihtelee, eikä sen käyttely sinänsä tee tutkimuksesta määrällistä tai edes erityisen empiiristä. Myös nojatuolissa istuskeleva, puhtaasti kieli-intuitionsa pohjalta tutkimusta tekevä voi hyödyntää korpusaineistoa. Usein nojatuolitutkijoiden (joita ovat muun muassa Chomsky, Talmy ja Langacker) työ kuitenkin kuvaa ”idealisoitua” kielentaitajan tietoa kielestä tai Lyonsin (1977: 622–635) luokittelu lainatakseni (riisuttuja) systeemilauseita (*system sentences*), vastakohtana (sotkuisille) tekstilauseille (*text sentences*), jotka ovat korpustutkimuksen kohteena.

Korpusaineistoa hyödyntävät tutkimukset voidaan jakaa (i) korpusesimerkein tuettui-

hin (*corpus-illustrated*), (ii) korpusvetoisiin (*corpus-driven*) ja (iii) korpuspohjaisiin (*corpus-based*) tutkimuksiin (Tognini-Bonelli 2001; Tummers yms. 2005: 234–238).

Jantusen (2009: 102–103; ks. myös Biber & Jones 2005: 152) mukaan korpusesimerkein tuetuissa (i) tutkimuksissa korpus toimii lähinnä materiaalipankkina, joka tuo introspektion, intuition ja laboratoriolauseiden rinnalle tietoa todellisesta kielenkäytöstä. Korpusvetoisessa tutkimuksessa (ii) korpuksen rooli on suurempi. Korpus ei ole vain aineistopankki, vaan tutkimusaineiston ensisijainen lähde, jota hyödynnetään olemassa olevien hypoteesien testaukseen ja tarkentamiseen. Korpuspohjaisessa tutkimuksessa (iii) sitoutuminen korpuksen on vahvinta: tutkimus tehdään korpusdatan pohjalta mahdollisimman vähin lähtöoletuksin, äärimmäisen aineistolähtöisesti. Korpuspohjaisen tutkimuksen tavoitteena ei ole pelkästään todistaa aiempia havaintoja ja teorioita oikeiksi, uudelleenmuokattaviksi tai vääriksi, vaan se pyrkii nostamaan esiin myös sellaisia kielen ilmiöitä, joita ei ole mahdollista havaita ilman laajan aineiston todistusvoimaa ja määrällistä, tilastollista tietoa. (Tognini-Bonelli 2001: 84; 2002: 75.)

Tutkimukseni on nähdäkseni lähimpänä korpuspohjaista (iii), sillä suhtaudun tutkimusongelmaani hyvin aineistolähtöisesti. En korpusaineiston lisäksi käytä muuta tekstiaineistoa¹⁶, paitsi sanakirjoja skeemojen määrittelyssä. Divjak ja Gries (2006, 2009) korostavat, että käyttäytymisprofileihin pohjautuva lähestymistapa on itse asiassa radikaalisti korpusvetoinen.

Seuraavaksi tarkastelen määrälliseen tutkimukseen liitettyjä periaatteellisia lähtökohtia.

2.1.2 Korpusaineiston keruu ja käsittely

Empiirisen kielentutkimuksen kaksi metodista perusongelmaa ovat kielenaineoksen valinta eli otoksen määrittäminen tutkittavasta perusjoukosta sekä otoksen koon määrittäminen. Edellisessä tapauksessa kyse on validiteetista, siitä, miten hyvin otos edustaa populaatiota. Tutkimuksessa voidaan tarkastella koko populaatiota eli mitata muuttujien arvo kaikista populaation yksiköistä, tai tutkimus voidaan rajata osaan populaatiosta ja yleistää tästä osasta havaitut tulokset koko populaatiota koskeviksi. Tutkimuksen korkea validiteetti tarkoittaa lisäksi, että ilmiötä kuvaamaan valitut muuttujat todella vastaavat kuvattavaa ilmiötä ja että kaikkien kyseessä olevien muuttujien kannalta tärkeät alakategoriat ovat tulleet huomioonotetuksi. (Luojola 2006: 17; Pitkänen & Kohonen 1984: 66–70.)

Otoksen koon määrittämisessä taas on kyse tutkimuksen reliabiliteetista. Optimaalinen otoskoko on sopiva: ei liian pieni jolloin tulokset eivät ole luotettavia, eikä liian suuri, jolloin

¹⁶ Poikkeuksena keksimäni tai muokkaamani lauseet.

olla tehty turhaa keruutyötä. Korkealla reliabiliteetilla tarkoitetaan käytetyn luokittelun (eli muuttujien) tarkkuutta ja onnistuneisuutta: sitä, että tutkimus ei sisällä ongelmallisia tai liian yksityiskohtaisia luokituksia (Karlsson yms. 1980: 91; Luoja 2006: 17; Pitkänen & Kohonen 1984: 66–70). Sivuan luokitusreliabiliteettiä¹⁷ luokkien koodausosiossa (luku 4)¹⁸.

Karlsson yms. (1980: 90–97) kuvaavat käsitteellä *jäätymispiste* otoskokoa, jonka yläpuolella merkitseviä muutoksia muuttujien alakategorioiden frekvensseissä ei enää todeta. Havainto on, että ainakin homogeenisten tekstilajien syntaktisten piirteiden erittelyssä jo suhteellisen pieni aineisto – heidän tapauksessaan muutaman sadan lauseen¹⁹ korpus – riittäisi useimpien alakategorioiden analyysiin ilman, että tulosten reliabiliteetti kärsii. Silti korpuksen kokoon liittyvien metodisten ongelmien pohdiskelu on oleellinen osa jokaista empiiristä tutkimusta (ks. Karlsson mts. 96). Salmela ja Kohonen (1977) huomauttavat, että ratkaisevimmin otoskoon vaikuttavat alakategorioiden lukumäärä sekä kunkin alakategorian esiintymien lukumäärä (frekvenssi). Isot alakategoriat edellyttävät isompaa otosta. Nyrkissännöksi Karlsson yms. (mts. 93) esittävät, että validisti poimittu otos tuottaa reliaabelin frekvenssijakauman, jos jokaisesta alakategoriasta on ainakin viisi esiintymää. Työssäni alakategorioita on maltillisesti. Lisäksi yhdistelen niitä, milloin tilastollisen luotettavuuden arviointi tätä vaatii ja yhdistely on teoreettisesti mielekästä, eivätkä yhdistelmäkategorioiden muodosta epäluonnollisia luokkia. Työni otoskoko onkin nähdäkseni riittävä²⁰.

Luolojan (2006: 11, 23–24) mukaan määrällisen työn aineiston laatimisessa keskeisiä periaatteita ovat lisäksi i) keruumenetelmien yhdenmukaisuus, ii) se, ettei aineistoa keräämisen jälkeen manipuloida esimerkiksi ongelmallisia tapauksia eliminoimalla sekä iii) se, että aineisto on poimittu satunnaisotannalla, jossa kullakin populaation yksiköllä on yhtä suuri todennäköisyys päätyä otosjoukkoon. Jo johdannossa kerroin soveltaneeni niin sanot-

¹⁷ Luokitusreliabiliteettia voidaan kohottaa selkeyttämällä kuvausmallia ja kriteeristöä, luopumalla ongelmallisista luokituksista (mikäli ne eivät ole keskeisiä), korvaamalla heikot luokitukset paremmilla ja yhdistelemällä pienempiä luokkia suuremmiksi (haittana informaatiokato) (Pitkänen & Kohonen 1984: 71). Divjakin ja Griesin (2006: 28) tapaan välttelyn tulkinnanvaraisia luokkia ja suosin ”objektiivisia” kielenpiirteitä mittaavia luokkia (esim. TEMPUS, ks. luku 4.3). Semanttisissa luokissa pyrin koodauskriteerien läpinäkyvyyteen.

¹⁸ Oikeastaan *ero*-nominin suhteen otannassa kyse on näytteestä (vrt. Luoja 2006: 25), sillä rajasin *ero*-nominin sijamuotoja pois kielitieteellisen harkintani mukaan.

¹⁹ Karlssonin yms. (1980: 93) sanatarkka päätelmä on: ”Niin kauan kuin alakategorioita on ’kohtuullisesti’, karkeasti sanottuna vähemmän kuin kymmenen, niiden frekvenssijakauma ei ole kovin vino (niin ettei mikään ole kovin harvinainen), ja otos on homogeeninen (eikä koostu toisistaan jyrkästi poikkeavista tekstilajeista), *muutaman sadan* lauseen otos on itse asiassa riittävä useimpien keskeisimpien muuttujien frekvenssijakauman luotettavaa selvittämistä varten.”

²⁰ Eräitä alakategorioita ja alamerkityksiä, kuten skeemaa EROTA.KONKRETIA (*Erosimme puistossa*), on otoksessani niukasti mahdollisesti tarkastelemani korpuksen rekisteristä johtuen (ks. luku 3). Harvinaisten kielenpiirteiden tai alamerkitysten lukumäärää voidaan lisätä kasvattamalla otosta tai sisällyttämällä otokseen aineksia monenlaista kieltä sisältävistä (korpus)aineistoista, mikäli sellaisia on saatavilla (esimerkiksi kaunokirjallisuutta). ”Täydellisempi” *erota*-peseyeen analyysi edellyttäisi tekstilajillisesti heterogeenisempää otosta.

tua yksinkertaista satunnaisotantaa (ks. esim. Holopainen & Pulkkinen 2003: 29–35) ainoastaan *ero*-nominiin, kun taas *erota-* ja *erottaa-*verbien HS.fi-populaatiosta tutkin kaikki esiintymät eli koko populaation. Keruumenetelmien valossa työni on yhdistelmä kokonaistutkimusta ja otostutkimusta, mutta kieliteellisesti aineistoni on otos: tarkastelen HS.fi-korpuksesta poimittua, 1666 tekstikatkelman joukkoa. Kielen tyhjentävä ”kokonaistutkimus” olisikin mahdotonta. Huomioitavaa on myös, etten yleistä *ero*-nominin satunnaisotoksesta havaitsemiani tuloksia koko *ero*-nominin populaatiota koskeviksi (kuten tyypillisesti tehdään), vaan suhtaudun *ero*-nominin otokseen ikään kuin se edustaisi samankaltaista kokonaistutkimusta kuin *erota-* ja *erottaa-*verbien äärelliset otokset. Näinhän ei ole. Tätä ratkaisua voidaan kritisoida puhtaasti tilastollisesta näkökulmasta, mutta kielitieteellisesti en näe siinä ongelmaa. Sopii kertoa, etten ole keräämis- ja koodaamisprosessissa manipuloinut aineistoani, paitsi tapauskohtaisesti rajoittamalla, miten paljon kontekstia *ero*-lekseemien ympäriltä analysoin. En ole korjannut kirjoitus- (tai ajatus-)virheitä. Annotaation olen tehnyt aitojen kielikatkelmien varassa; käyttämiäni korpuskatkelmia olen kuitenkin lyhennellyt ja korjannut oikeinkirjoitusta, milloin se häiritsee luettavuutta.

Semanttisen painotukseni takia työ ei ole kvantitatiivisuudessaan ”luonnontieteellisen eksakti”. Vaikka tutkimustulokseni kumpuavat määrällisten analyysivälineiden tuloksista, päättelyvälineenä käytän (kieli-)intuitiotani. Suodatan tilastollisten testien tulokset päättelyni kautta, ja teen niistä omat arvioni. Semanttinen analyysi lieneekin mahdotonta ilman kielentutkijan intuition osallisuutta (ks. Huomo & Sivonen 2010: 33–36). Erilaisten näkökulmien yhdistely sopii hyvin kognitiivisen kielentutkimukseen, jonka piirissä on kritisoitu kielitieteeseen kehittyneitä tiukkoja dikotomioita (Leino 1993: 34). Jakoa määrälliseen ja laadulliseen tutkimukseen voitaneen pitää yhtenä sellaisena. Määrällisissä töissä esiintyy aina laadullisia työvaiheita, joita ovat esimerkiksi analyysityökalujen ja teorian valinta sekä tulosten raportointi. Määrällinen tutkimus myös edellyttää laadullisesti perusteltuja tulkin-toja, jotka pohjautuvat tietoon kielellisten piirteiden jakaumasta (Pitkänen & Kohonen 1984: 10)²¹. Laadullisen ja määrällisen tutkimuksen erottaminen onkin ehkä keinotekoista; tutkimusstrategioiden yhdistämisessä on etuna, että saadaan esiin molempien vahvoja puolia.

Erityisesti muuttujien kategorioiden luominen ja aineistoesimerkkien merkitseminen on tulkinnan sisältävää ja subjektiivisen intuition ohjaamaa, ei absoluuttista toimintaa (ks.

²¹ Pitkänen ja Kohonen (1984: 11) pitävät intuitiopohjaista ja korpuslähtöistä tutkimusta toisiaan täydentävinä tutkimusstrategioina: korpus tutkimuksessa etsitään (laadullisiin) kysymyksiin määrällisiä vastauksia, ja korpus tuo aina väistämättä esille uusia kysymyksiä, jotka vaativat analyysiä ja suhteuttamista lähtökohdaksi olleisiin oletuksiin ja teoreettisiin lähtökohtiin.

Arppe 2008: 30 ja siellä mainitut lähteet). Koska käyttämäni HS.fi-korpuksen tekstiesiintymiin ei ole merkitty minkäänlaisia morfosyntaktisia saati semanttisia tunnisteita, muuttujien arvot olen koodannut manuaalisesti ja R-kielelle kirjoittamalla skripteillä. Manuaalinen koodaustyö on hidasta ja sisältää kosolti saumaa virhelähteille²². Olen koodannut ja tarkistanut työni yksin, mikä poikkeaa esimerkiksi Karlssonin yms. (1980: 8–9) ja Arppen (2008) ”intersubjektivisesti” koodatuista, joskin yksittäisten tutkijoiden hallinnoimista töistä. Haitana on, että jotkin luokitukset saattavat olla subjektiivisesti vinoutuneita ja etuna, että yhden koodaajan suorittama analyysi sentään usein on johdonmukaisempi (ks. Pitkänen & Kohonen 1984: 71). Koodausvirheitä olen poistanut ristiintaulukoimalla muuttujia, jotka sisältävät mahdottomia tai toisensa poissulkevia luokituksia. Esimerkiksi agenttia mittaavien muuttujien X_SEM ja X_MUOTO ristiintaulukointi paljastaa, onko NA:ksi²³ koodatuille eli ilmipanemattomille agenteille (X_MUOTO) vahingossa koodattu semanttinen arvo (X_SEM). Muuttujien PÄÄLUOKKA ja X_MUOTO ristiintaulukointi taas paljastaa, onko passiivilauseeseen koodattu läsnä oleva agentti, kun taas muuttujien Z_MUOTO ja RAKENNE ristiintaulukointi paljastaa, onko luokkaan VAIN_Y kuuluvalla havainnolle (*erottaa*-verbissä argumenttirakenne [S]VO) vahingossa koodattu Z-argumentti, lähde. Lopuksi olen manuaalisesti tarkastanut yksittäisiin luokkiin (kuten Y_SEM_JOUKKO) koodatut havainnot. Mikäli olen havainnut ristiriitaisia aineksia, kuten sanan *koulu* luokassa Y_SEM_JOUKKO, olen vaihtanut sen luokaksi Y_SEM_INSTITUUTIO. En ole pitänyt kirjaa sadoista tekemistäni ristiintaulukoinneista ja luokkien tarkastuksista. Viimekädessä koodauksen virheettömyys ja koodausratkaisujen oikeellisuus on yksin päättelyni, tarkkuuteni ja johdonmukaisuuteni varassa.

Annotoinnin lisäksi tutkimukseni laadullis-intuitiivisiin vaiheisiin kuuluu monimerkityksisyyden jakaminen skeemoiksi. Tämän työvaiheen esittelen luvussa 3. Työvaihe pohjaa tutkimuksen merkityskäsitykseen, josta teen selkoa seuraavaksi.

2.2 Työn merkityskäsitys

Semanttisen analyysin kannalta on ratkaisevaa, millainen merkitysnäkemyksellä tutkijalla on. Merkitysnäkemykseni nousee funktionaalisen ja kognitiivisen kielentutkimuksen perin-

²² Annotoimattomia (isoja) aineistoja varten on olemassa myös suomeen sovellettavia automaattisia työkaluja. Manuaalinen annotointi on kuitenkin edelleen niin sanottu *gold standard* (Evert ja Krenn 2001).

²³ Notaatio NA tarkoittaa alakategoriaa, jossa tiettyä kielenpiirrettä ei esiinny. Kyse ei ole tyhjästä solusta. Esimerkiksi RESIPROOKKI sisältää luokat LÄSNÄ (piirre esiintyy) ja NA (piirrettä ei esiinny lauseessa).

teestä vastakohtana strukturalistiselle ja generatiiviselle perinteelle. Kognitiivinen kielentutkimus lähtee siitä, että kieltä ei tarkastella irrallisena lingvistisenä komponenttina tai autonomisena ja itseriittoisena järjestelmänä, vaan kytköksessä ihmisen kognitioon ja kieliyhteisössä tapahtuvaan vuorovaikutukseen (Geeraerts 2006: 4; Onikki-Rantajääskö 2001: 30; Leino 1993: 53)²⁴.

Esimerkiksi Langacker (1987: 6–7, 13, 101) korostaa, että kielessä heijastuu samantyyppiset perusprosessit kuin muussakin kognitiossa, kuten samankaltaisuuksien ja erojen havaitseminen sekä hahmon ja taustan erottaminen. Kielen merkitysrakenteiden ajatellaan motivoituvan sitä kautta, että kieltä käyttää ihminen, joka on kognitionsa, aistiensa ja kulttuurisesti jäsentyneen ruumiillisuutensa välityksellä jatkuvassa vuorovaikutuksessa ympäristönsä kanssa (Johnson 1987: xvi; Onikki-Rantajääskö 2001: 31 ja siellä mainitut lähteet). Näin ymmärrettynä kieli on intersubjektiiivinen, dynaaminen ja käyttöpohjainen ilmiö (Langacker 1987: 62–67). Kun kieli ymmärretään käyttöpohjaiseksi, myös kielentutkimuksen sopisi olla käyttöpohjaista: käyttäytymisprofileihin pohjautuva määrällinen lähestymistapa onkin sitä mitä suurimmissa määrin (vrt. Gries 2006, 2012).

2.2.1 Kategorisointi

Kognitiiviselle kielentutkimukselle – joka on pikemminkin keskustelufoorumi kuin yhtenäisen tutkimussuuntaus (Onikki-Rantajääskö 2001: 30) – on ominaista, että se hylkää syntaksin autonomian ja tarkastelee kielen yksiköitä (lekseemejä ja rakenteita) merkityksestä käsin (Onikki-Rantajääskö mts. 47). Merkitys ymmärretään mentaaliseksi ilmiöksi, käsitteistykseksi (mm. Leino 1993: 55). Taustalla on Langackerin (1991: 76–77) kognitiivisen kielioopin näkemys siitä, että kieli koostuu konventionaalisista symbolisista yksiköistä, joilla on sekä semanttinen että fonologinen napa – sekä merkitys että muoto. Merkitystä edustaa predikaatti, yksikön semanttinen napa.

Sanoja ei nähdä staattisina, täsmällisen merkityksen sisältävinä säiliöinä, eikä kielelliseen merkkiin yhdistyvää ”ensyklopedista” tietoa eroteta ”lingvistisestä” merkityksestä (Leino 1993: 30–39; 101). Kielellinen ilmaus nähdään pikemminkin vihjeeksi, joka käynnistää tulkitsijassa päättelyketjun, jossa vuorovaikutusprosessilla ja kontekstilla on merkittävä tulkintaa ohjaava rooli (Leino mts. 44–46, 73; Onikki-Rantajääskö mts. 31.) Seurauksena on ihanne ekonominen kuvauksen sijaan mahdollisimman hienojakoisesta mallinnuksesta (mm. Leino mts. 133; Langacker 1987).

²⁴ Onikki-Rantajääskö (2001: 30) huomauttaa emergenssistä: olisi metateoreettista naivismia olettaa, että kielen käsitejärjestelmä palautuisi suoraan kognition prosesseihin.

On syytä korostaa, että tutkimukseni määrällisen otteen takia (ks. luvut 2.3 ja 4) muutajat ja niiden alakategoriat (eli tarkkailtavat kielenpiirteet) on jouduttu jakamaan toisensa poissulkeviin luokkiin. Se presupponoi, että analysoitava ilmiö on siinä määrin epäjatkuva ja ”luokkahakuinen”, että siitä voidaan lohkoa erillisiä jaksoja tai osia, joille voidaan yksikäsitteisesti määrittää mittausarvoja. Näin ei tietysti aina ole. Tällainen luokittelu on päinvastoin osittain ristiriidassa esittelemäni kognitiivisen merkitysnäkemyksen kanssa (vrt. Leino 1993: 133). Langackeriin (1987), Lakoffiin (1987) ja Lakoffiin ja Johnsoniin (1980) pohjautuvassa kognitiivisessa perinteessä on ohjelmallisesti vastustettu jyrkkärajaisia ”klassisia kategorioita”, joissa kategoriaan kuulumisen luokkajäsenyys ei ole asteittaista vaan ehdotonta siten, että jokainen entiteetti kuuluu kiistatta joko luokkaan tai sen ulkopuolelle. Klassisille kategorioille on Leinon (mts. 33) mukaan ominaista, että ne määräytyvät välttämättömien ja riittävien ehtojen nojalla binäärisistä merkityskomponenteista (esim. *naisen* merkitys sisältää piirteet +IHMINEN, +AIKUINEN ja +NAISPUOLINEN).

Vaihtoehtoksi on nostettu Roschin (1978) kokeisiin pohjautuvaa prototyypiteoriaa, joka korostaa luokkien sumearajaisuutta, luokkajäsenyyden asteittaisuutta ja luokittelun joustavuutta (ks. Leino 1993: 38). Wittgensteinin mukaan (1981: 71) ”sanalla on kokonainen merkityksien perhe”, jonka jäsenten välisiä suhteita hän luonnehtii perheyhtäläisyydeksi. Prototyypiteorian eetos on, että kielenkäytön esiintymät voivat olla enemmän tai vähemmän edustavia ja siten sopivia jonkin kategorian jäseneksi. Prototyypin yksi määritelmä on, että se edustaa luokan mallijäsentä, sitä jonka nojalla yksittäinen merkitysesiintymä liitetään kategoriaan; prototyyppi ilmentääkin usein skeemaa, joka sanktioi mallijäsenen tai -jäseniä (Onikki 1994: 117; Onikki-Rantajääskö 2001: 33; Leino 1993: 36–39). Periaatteen tasolla työssäni on vain prototyyppisiä kategorioita, vaikka kvantitatiiviset luokittelut eivät tuokaan sitä esille. Prototyyppisyys näkyy kuitenkin kategorioiden kriteeristöissä (ks. luvut 3.3 ja 4).

2.2.3 Monimerkityksisyys ja sen kuvaaminen

Tutkimani lekseemit *ero*, *erota* ja *erottaa* ovat oletukseni mukaan polyseemisiä eli monimerkityksisiä: niissä samaan fonologiseen rakenteeseen on liittynyt useampia semanttisia rakenteita tai predikaatteja (ks. Leino 1993: 73, 162), joita kuvaan skeemojen avulla. *ero*-lekseemeillä tuotetaan monia merkityksiä²⁵.

Tuggy (2006: 167–169) jakaa monimerkityksisyyden monitulkintaisuudeksi (*am-*

²⁵ Se, milloin tietyn sanan käyttö edustaa samaa ja milloin eri ”merkitystä”, ei ole kielenkäytön ongelma vaan teorian. Äidinkielen puhuja navigoi monimerkityksisten lekseemien viidakossa yleensä kuin kotonaan.

biguity) ja merkityksen sumeudeksi (*vagueness*). Leinon (1993: 162) mukaan monitulkintaisuudesta on kysymys silloin kun, semanttiseen rakenteeseen liittyy kaksi itsenäistä, toisistaan selvästi erottuvaa merkitystä, joihin ei liity vakiintunutta ja elaboraatioetäisyydeltään läheistä ne molemmat kattavaa skeemaa. Elaboraatioetäisyys on Langackerin (1987) käsite, joka kuvastaa skeeman ja sen toteutuman välistä (semanttista) etäisyyttä: mitä vähemmän lisäspesifikaatiota toteutuma tuo skeemaan, sitä lyhempi niiden elaboraatioetäisyys on (Leino mts. 161).

Esimerkkinä monitulkintaisuudesta Leino (mts. 162) antaa suomen kielen sanan *susi*, jonka merkitykset 'canis lupus' (kuviossa 1e laatikko A, ks. s. 18) ja 'työn epäonnistunut tulo' (kuviossa 1e laatikko B) ovat molemmat vakiintuneita, mutta ne yhdistävän skeeman asema verkossa on hyvin heikko (merkityksen verkkoluonteisuudesta ks. luku 7). *ero*-pesyeessä monitulkintaisena voidaan nähdä ainakin *erota*-verbin merkitykset 'erota työpaikasta' ja 'olla erilainen ominaisuudeltaan ja laadultaan'. Molemmat merkitykset ovat vakiintuneita, mutta nykykielenkäyttäjän on nähdäkseni hankala keksiä niitä yhdistävää yläskeemaa.

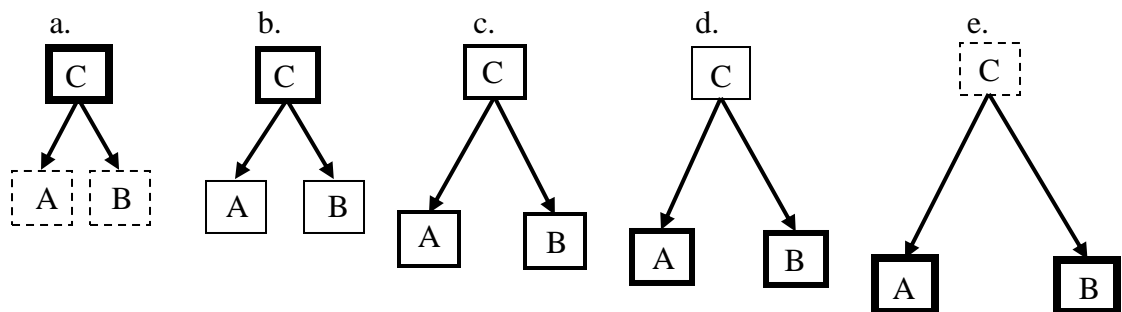
Merkityksen sumeudesta tai joustomerkityksestä on Leinon mukaan kyse silloin, kun spesifisempiä merkityksiä yhdistää verraten vakiintunut ja elaboratiivisesti läheinen skeema. Esimerkkinä hän antaa sanan *täti*: suomen kieli tekee eron äidin ja isän veljen välillä (*eno* : *setä*), mutta ei tee eroa 'äidin sisar' ja 'isän sisar'. Niitä yhdistävä skemaattisempi predikaatti on Leinon mukaan 'vanhemman sisar' (kuviossa 1a laatikko C). Se on suhteellisen korostettu ja sen elaboraatioetäisyys predikaatteihin 'äidin sisar' (kuviossa 1a laatikko A) ja 'isän sisar' (kuviossa 1a laatikko B) on vähäinen. Merkityksen sumeus tarkoittaa siis sitä, että samaan fonologiseen rakenteeseen liittyy kaksi semanttista rakennetta, jotka molemmat ovat johdettavissa läheisestä yläskeemasta kontekstin avulla (ks. Evans & Green 2006: 340). Merkitysvaihtelussa kontekstilla onkin – kuten Tuggy (2006: 177) ja Leino (1993: 163) korostavat – ensiarvoinen asema, sillä monimerkityksisen kielenyksikön eri merkitykset aktivoituvat paitsi tietyn merkitysvariantin vakiintuneisuuden, erityisesti kontekstin vihjeiden avulla. Työni onkin pitkälti tällaisten kontekstin vihjeiden analyysiä.

ero-pesyeestä merkityksen sumeutta edustaa kandidaatin tutkielman havaintojeni pohjalta esimerkiksi *erota*-verbin merkitykset 'erota virasta' ja 'erota työpaikasta', joita yhdistää skemaattisempi predikaatti 'erota jonkin yhteisön tms. yhteydestä'. Merkitys syntyy kontekstin vihjeiden avulla; sanan merkitys joustaa kontekstin mukaan.

Monitulkintaisuuden ja merkityksen sumeuden lisäksi Leino (1993) erottaa kolmanneksi monimerkityksisyyden tyypiksi polysemian (kuvio 1c). Käytän merkityksenvaihtelun kattokäsitteenä monimerkityksisyyden käsitettä. Se kattaa käytössäni monitulkintaisuuden,

merkityksen sumeuden sekä polysemian. Polysemia-käsitteen varaan sumeuden vastapariksi jatkumolla sumeus : polysemia (*vagueness* : *polysemy*). Leinon erottelussa polysemia sijoittuu jatkumolla sumeuden ja monitulkintaisuuden välille siten, että eri merkitysten välillä on semanttinen linkki, mutta merkitykset ovat sen verran itsenäisiä, että on mielekästä puhua eri merkityksistä. Esimerkkinä polysemiasta Leino esittää sanan *piippu*. Sen merkitykset 'tupakan polttamiseen käytetty väline', 'savupiippu' ja 'tuliaseen osa' ovat erillisiä mutta sitenkin toisiinsa liittyneitä. Leinon mukaan yhteistä erilaisille piipuille on, että niistä tulee savua ja että ne ovat kaikki jossain määrin putkimaisia.

Tuggyta (2007) peräisin olevassa kuviossa 1 C-kirjaimin varustetut laatikot esittävät merkityksen yläskeemaa ja A- ja B-kirjaimin varustetut laatikot merkitysvariantteja. Laatikoiden reunojen viivatyypit ilmoittavat kategorioiden korostuneisuusastetta (*salience among the cohering structures*, Tuggy mts. 98) tai konventionaalistumisen astetta muistisysteemissä (*degrees of entrenchment in memory*, Evans & Green 2006: 340). Katkoviiva ilmoittaa, etteivät merkitykset A ja B ole varastoituna kielenkäyttäjän semanttisessa muistissa erillisinä yksiköinä, vaan ne tulkittuvat kielenkäyttötilanteessa (*on-line processing*, Evans & Green mts. 341). Nähdäkseni laatikoiden välinen etäisyys ilmoittaa pystysuunnassa ylä- ja alakategorioiden välistä elaboraatioetäisyyttä ja vaakasuunnassa kategorioiden välistä semanttista läheisyyttä (eli sitä, miten paljon perheyhtäläisyyttä kategoriolla on).



Kuvio 1. Merkityksen sumeus (a), polysemia (c) ja monitulkintaisuus eli homonymia (e) liukumolla Tuggyn (2007: 99–100; 2006: 177; ks. myös Leino 1993: 162–163) mukaan.

On syytä korostaa, etteivät monitulkintaisuus (homonymia) ja merkityksen sumeus (ja polysemia) ole jyrkkärajaisia ja dikotomisia kategorioita. Tuggyn (mts. 99) mukaan kategoriat ovat pikemminkin sumearajaisia ja sijoittuvat jatkumolle, joka ulottuu monitulkintaisuuden (useampia erillisiä, epäyhtenäisiä merkityksiä; kuviossa 1e) ja polysemian (erillisiä mutta yhteneviä merkityksiä; kuviossa 1c) kautta merkityksen sumeuteen (yksi yhtenäinen merkitys; kuviossa 1a).

Lisäksi on syytä vielä painottaa, etten sitoudu työssäni polysemiaikäsiytykseen, jossa

sanan alamerkitykset nähdään toisistaan jyrkästi erillisinä. Semantiikassa ei ole Onikki-Rantajääskön (2001: 215 ja siinä mainitut lähteet) mukaan onnistuttu esittämään yhtäpitäviä kriteereitä, joiden mukaan jollekin ilmauksen käytölle annetaan alamerkityksen status (vrt. luku 3.3.). Uskon pikemminkin, että merkitys hajautuu syntagman eri elementtien kesken (vrt. *distributed meaning*, Sinha & Kuteva 1995). Tällaisessa näkemyksessä merkitys rakentuu niin sanoakseni synergisesti kollikaation jäsenten yhteisvaikutuksesta, jolloin on keinote-koista puhua yksittäisen lekseemin ”merkityksestä”. Alamerkityksistä ja skeemoista puhu-minen edustaa työssänikin kuvauskäytännön tarpeisiin sopivaa välttämätöntä karkeistusta.

2.2.2 Skeemat ja prototyypisuus

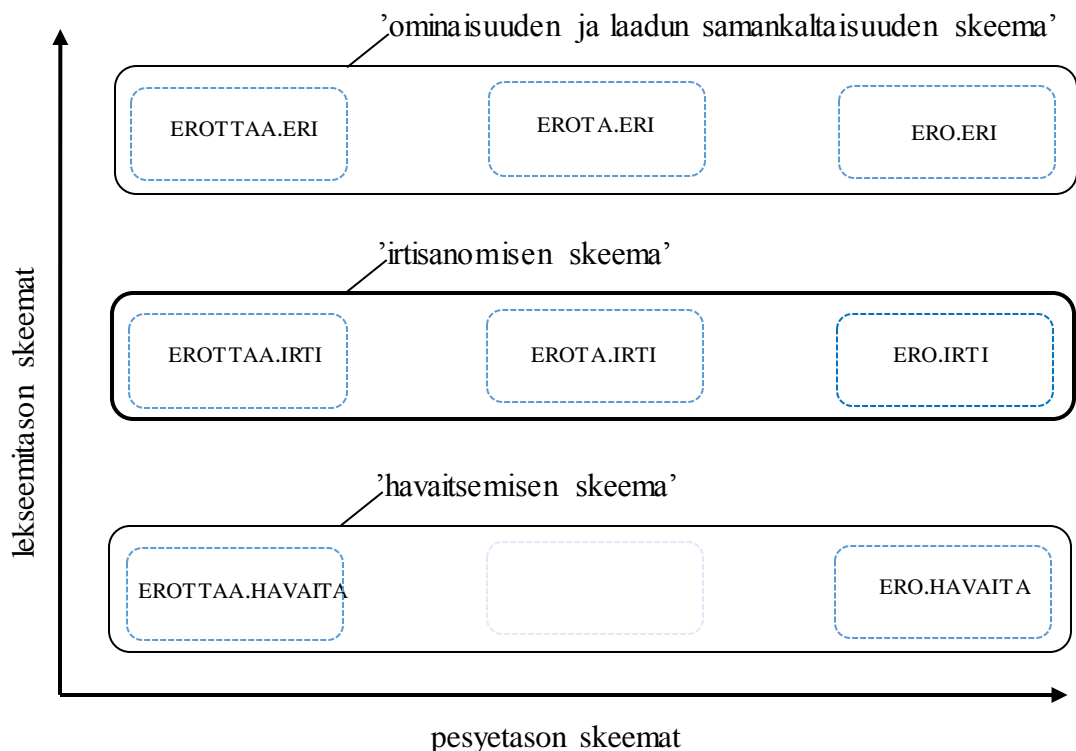
Kognitiivisessa traditiossa merkitystä on kuvattu skeemojen avulla (Leino 1993: 161–164). Skeema on psykologiasta peritty, useisiin tieteisiin sovellettu käsite, jolla on kielitieteen si-sälläkin erilaisia määritelmiä. Langackerin kognitiivisessa kieliopissa (1987, 2008) skeema on yksittäisistä merkitysesiintymistä tehty yleistys, säännön vastine, joka sulkee sisäänsä kaikki määrittämänsä luokan jäsenet (Onikki-Rantajääskö 2001: 32; Onikki 1994: 115–116; Leino 1993: 107–108). Skeema kuvaa siis karkeapiirteisesti sen, minkä skeeman toteuma –aktuaalinen esiintymä kielenkäytössä – kattaa (Onikki 1994: 115; Leino mts. 64). Langackerilla skeema käsittää paitsi kielellisen rakenteen yleistyksen myös käsitteistyksen abstrak-tiotason (Onikki mts. 126).

Prototyyppi ja skeema ovat, kuten Onikki-Rantajääskö (2001: 31–33) huomauttaa, kuitenkin hankalasti kuvattavia, eikä prototyyppikuvaus usein juuri eroa luokan tärkeimpiä piirteitä erittelevästä ekonomisesta kuvaustraditiosta. Tämä tutkimukseni ei tee poikkeusta.

Tutkimuksessani käytän käsitettä skeema, mutta määrittelen sen omalla tavallani. Skeema tarkoittaa työssäni *ero*-lekseemien merkitysrakenteesta kandidaatin tutkimukseni pohjalta tekemiäni havaintojen, intuition ja sanakirjojen avulla abstrahoimiani yleistyksiä. Hyödynnän skeema-käsitettä ensinnäkin eritellessäni *ero*-lekseemien vakiintuneita käyttö-tapoja, ”pesyeen merkitystyyppejä”, joista monilla on edustuma kaikkien *ero*-lekseemien merkitysrakenteessa. Kutsun näitä skeemoja pesyetason skeemoiksi. Pesyetason skeemat kytkevät *ero*-lekseemien merkitysvariantteja (skeemoja) sarjoiksi. Tutkimuksessani tarkoi-tukseni on osoittaa, että saman pesyeskeeman toteutumia yhdistää samankaltainen semant-tinen ja myös morfosyntaktinen käyttäytymisen, analogia käyttäytymisprofiilien tasolla (ks. esim. s. 171). Pesyeskeema on tutkimuksessani tämän semanttisen ja morfosyntaktisen tie-don sisältävänä tietorakenne, siis eräänlainen lekseemitasosta irrallinen konstruktio (vrt. Goldberg 1995: 1, 6–7; ks. myös Onikki-Rantajääskö 2001: 184–185).

Lisäksi viitataan skeemalla *ero*-lekseemien merkitysvariantteihin (ks. luku 3). Skeema tarkoittaa tällöin tietynlaisessa syntagmassa tai kollikaatiossa esiintyvän *ero*-”semeemin” skemaattista yleistystä (vrt. skeema Langackerilla 1987). Kutsun näitä skeemoja myös lekseemitason skeemoiksi. Ne jakavat *ero*-lekseemien merkitysrakennetta erilaisiksi alakategorioiksi (ks. Leino 1993: 161). En kuitenkaan väitä, että *ero*-lekseemien merkitysrakenne jakautuisi alakategorioiksi esittämieni skeemojen mukaisesti. Merkitysrakenteen silpominen skeemoiksi johtuu pikemminkin käytännön analyysi- ja kuvaustarpeista.

Koska skeema sisältää yksilöllisen semanttis-syntaktisen profiilin lisäksi tietorakenteeseen kytkeytyvän ensyklopedisen taustatiedon, sitä ei ole mahdollista sanallistaa tyhjentävästi (taustatiedon kuvausongelmista ks. Onikki-Rantajääskö 2001: 31). Ratkaisuni on viitata skeemoihin leksikografian tavoin parafraaseilla, jotka nostavat esiin skeeman ilmeisimmän (protyyppisimmän) ominaisuuden. Oletettuja pesyeskeemoja ovat muun muassa ’irtisanomisen skeema’ ja ’ominaisuuden ja laadun samankaltaisuuden skeema’. *erottaa*-verbin lekseemiskeemoja taas ovat esimerkiksi ’havaitsemisen skeema’ ja ’irtisanomisen skeema’. Skeemoja havainnollistaa kuvio 2.



Kuvio 2. Lekseemitason skeemat suhteessa pesyetason skeemoihin.

Kuviossa 2 katkoviivaiset laatikot ilmentävät lekseemitasolla abstrahoidut skeemoja eli alamerkityksiä ja mustareunaiset laatikot pesyetason skeemoja. Notaatiot, kuten

EROTTAA.HAVAITA ja ERO.HAVAITA, ovat lekseemitason skeemojen nimityksiä. Ne koostuvat kahdesta osasta. Alkuosa kertoo, mistä lekseemistä on kyse (EROTTAA, EROTA vai ERO). Loppuosa nimeää pesyeskeeman tunnistein (HAVAITA). Kunkin *ero*-johdoksen skemaattinen merkitysrakenne (tai ”ilmaisupotentia”) jäsentyy yksilöllisesti. Nuolet vihjaavat, että lekseemitason skeemat ovat assosiatiivisessa suhteessa toisiinsa sekä vaaka- että pystysuunnassa. Pystysuunnassa lekseemiskeemoja yhdistää johto-opillinen status eli se, mistä *ero*-lekseemistä on kyse. Keskeistä on se, että lekseemiskeemoilla tuotetaan ”eri merkityksiä”. Vaaka-suunnassa lekseemiskeemoja yhdistää pesyeskeema. Assosiatiivisuus ilmenee tällöin siten, että saman pesyeskeeman sisällä tuotetaan ”samankaltaisia merkityksiä”. Assosiatiivisuutta on toisin sanoen kahta lajia: johdosten sisäistä (esim. EROTTAA.HAVAITA : EROTTAA.IRTI) ja johdosten välistä (EROTTAA.IRTI : EROTA.IRTI).

Skeemojen muodostama teoreettinen merkitysjatkumo voidaan nähdä hierarkkiseksi siten, että ylimmällä tasolla sijaitsevat pesyetason skeemat, joiden toteutumia lekseemitason skeemat ovat. Pesyeskeemat ovat skemaattisia suhteessa lekseemiskeemojen spesifioimaan informaatioon siitä, miten toiminta ja sen osallistujat jäsennetään. Esimerkiksi *erottaa*-verbi tuo näyttämölle voimadynaamisen alkuunpanijan, joka puuttuu *erota*-verbin ilmoittamista tilanteista (ks. tarkemmin esim. s. 52). Silti sekä EROTTAA.IRTI että EROTA.IRTI ovat saman pesyeskeeman, ’irtisanomisen skeeman’, toteutumia. Varsinaiset kielenesiintymät ovat lekseemitason skeemojen toteutumia eli spesifikaatioita. Skeemat ovat karkeistettuja hahmoja, joita aineistoesimerkit tarkentavat. Pystysuunnassa skeemoja yhdistää skemaattisuus/to-teuma-suhde (ks. Leino 1993: 135). Prototyypisyys ilmenee siinä, että kaikki skeeman EROTTAA.IRTI korpustoteutumat eivät ole yhtä edustavia. Myös muuttujat ovat prototapauksen ympärille rakentuvia kategorioita. Malliin voidaan lisäksi teoretisoida ns. ylimmän tason superskeema (ks. Leino mts. 119), jolle on ominaista heikko vakiintuneisuus ja kognitiivinen korostumattomuus. Jatkossa käytän käsitettä kategoria sekä skeemojen että muuttujien yhteisnimityksenä. Muuttujien alakategorioita nimitän luokiksi.

Laatikkomainen esitystapa saattaa harhauttaa ajattelemaan skeemat jonkinmoiseksi erillisiksi merkityssäiliöiksi (ks. Leino 1993: 163). Skeemojen olemus on kuitenkin toinen: ne muodostavat yhtenäisen jatkumon, jossa kategoriat limittyvät ja päällekkäistyvät. Skeemojen luokkajäsenyys on asteittaisia ja noudattavat luokille ominaista prototyypisyysilmiötä. Leinon sanoin ilmaistuna (mts. 213–214) kyse on siitä, että skeeman täysin sanktioidut toteumat ovat sen edustaman luokan prototyyppisempiä edustajia kuin vain osittain sanktioidut. Mutta heikosti sanktioitu toteutumakin, siis kielenesiintymä jolla on vain vähän

luokan ominaisuuksia, kuuluu sekin luokkaan. Koska prototyyppisyyttä on hankala mallintaa määrällisesti, työni ei tee eroa perifeeristen ja prototyyppisten edustumien välillä. Kvantitatiivisin menetelmin on kuitenkin selvitettävissä (ks. luvut 5 ja 6), mitkä kielenpiirteet (muuttujat) ovat tyypillisiä kullekin skeemalle (tai klusterille).

Tutustun seuraavaksi siihen, mitä monimerkityksisyys on ja miten sitä voidaan kuvata.

2.3 Taustaa muuttujien koodaukselle

Käyn työssäni käytetyt muuttujat ja niiden koodauskriteerit yksityiskohtaisesti läpi luvussa 4. Ennen sitä on kuitenkin syytä tarkastella muuttujien teoreettista taustaa.

Jotta määrällisiä menetelmiä voisi soveltaa laadulliseen aineistoon – kieleen ja sen rakenteisiin – on kieli muunnettava laskutoimituksia sallivaan mitattavaan muotoon. Laadulliseen aineistoon, jossa huomiona ja ainoana mitattavana suurena on tutkittavan kohteen laatu tai ominaisuus, ei sen suuruus, paremmuus tai järjestys, voidaan soveltaa käytännössä ainoastaan nominaaliasteikollisia muuttujia (ks. Nummenmaa 2004: 33–35; Luojola 2006: 18). Nominaaliasteikolliselle muuttujalla²⁶ ominaista on, että sen koodaama havainto voi kuulua kerrallaan vain yhteen luokkaan ja että kaikki havainnot kuuluvat johonkin luokkaan. Luokat ovat siis ”tyhjentäviä” ja toisensa poissulkevia eikä niillä ole keskinäistä järjestystä. (ks. Karlsson yms. 1980; Luojola mts. 17–19.) Esimerkiksi *ero*-verbi voi olla joko aktiivi- tai passiivimuotoinen, ei koskaan molempia yhtäaikaaisesti.

Työssäni on kolmenlaisia muuttujia: semanttisia, morfologisia ja syntaksia. Pääpaino on *erottaa*- ja *erota*-verbien ydinargumenteissa – subjektin, objektin ja rektioadverbiaalin semanttisissa ja morfologisissa luokituksissa, tai yksinkertaisimmin: *ero*-lekseemien täydennyksissä. Täydennys ei ole pelkästään verbin ominaisuus, vaan sekä nominilla että verbillä voi Vilkun mukaan (ks. esim. 1996: 347; ks. myös VISK § 446) olla valenssimukaisia jäseniä eli täydennyksiä, jotka ovat pakollisia käsitteellisesti, toisin sanoen sanan merkityksen edellyttämiä, mutta eivät aina välttämättömiä lauserakenteessa. Perinteisesti puhutaan syntaktisesta valenssista, siitä, kuinka monipaikkaisia verbit ovat argumenttiensa suhteen (ks. Pajunen 1988: 15–19). Kognitiivinen kielentutkimus kuitenkin korostaa sitä seikkaa,

²⁶ Muuttujat jaotellaan usein kvalitatiivisiin (laadullinen) ja kvantitatiivisiin (määrällinen) muuttujiin, joista edeltäviä mitataan ainoastaan nominaaliasteikolla (laatuero- tai luokitteluasteikolla), kun taas jälkimmäisiä voidaan mitata järjestys-, välimatka-, suhde- tai absoluuttisella asteikolla. Laatueroasteikkoa pidetään tarkkuudeltaan vaatimattomimpana mitta-asteikkona, koska sen avulla voidaan laskea vain harvoja tilastollisia tunnuslukuja (laskutoimitukset eivät ole mielekkäitä ilman järjestyksen käsitettä). (Nummenmaa 2004: 30–37.)

että valenssi on pohjimmiltaan semanttinen ilmiö, joka ei rajoitu verbin paikkalukuun (Langacker 1987: 142, 277–327). Puhutaan semanttisesta valenssista, joka tarkoittaa, että yhdistyvien yksiköiden merkitysrakenteessa on korosteisia osarakenteita, joiden on sovittava yhteen ja jotka ovat osittain päällekkäisiä siten, että dependentin yksikön, kuten verbin, merkitysrakenteessa on paikkoja, joita autonomiset yksiköt, kuten substantiivit, tarkentavat (ks. Onikki-Rantajääskö 2001: 32; Langacker 1987: 82; Leino 1993: 91–93;). Onikki-Rantajääskön (mts. 32) mukaan semanttisuus näkyy juuri siinä, että osarakenteet, jotka toisinaan ovat ilmiöidä, voivat toisinaan jäädä vaille ilmiöidä mutta merkitysrakenteen edellyttämiksi. Ilmiö on vahvasti esillä työssäni.

Valenssinmukaisten jäsenten määrän perusteella *erota*-verbi on tyypillisesti kaksipaikkainen ja *erottaa*-verbi kolmipaikkainen. *erota*-verbiä esiintyy kuitenkin myös kolmipaikkaisena, kuten lauseessa *Vain Vasemmalta* [1] *erosi kaksi kansanedustajaa* [2] *omaksi ryhmäkseen* [3] *jo hallitusneuvotteluissa* (K843), jossa hakasuluissa olevat numerot demonstroivat paikkalukua. Tulostajasta lauseketta voidaan pitää täydennyksenä, koska lauseen merkitys olennaisesti muuttuu ilman sitä. *erottaa*-verbi taas esiintyy nelipaikkaisena, kuten esimerkissä *Yhdestä-kolmesta alueen partiosta* [1] *on hankala* [2 (nolla)] *erottaa miestä* [3] *liikenteenohjaukseen* [4] (K45)²⁷.

Niin kuin Divjak ja Gries (2006: 30) painottavat, käyttäytymisprofileihin pohjautuvassa lähestymistavassa on tärkeää koodata koko aineisto samojen parametrien mukaan. Työni suunnittelussa iso haaste onkin muuttujien sovittaminen sellaisiksi, että on mahdollista suorittaa tilastollista analyysiä *erottaa*-verbin muodostamien transitivilauseiden, *erota*-verbin muodostamien intransitivilauseiden ja *ero*-nominin sisältävien kopula-, intransitiivi- ja transitivilauseiden välillä. Tässä törmätään derivaation perusilmiöön (tai -ongelmaan): saman pesyeen eri johdoksilla koodataan usein samankaltaisia kielenulkoisia olioita ja tilanteita mutta ”eri näkökulmista”. Eri johdokset korostavat tai taka-alaistavat tilanteeseen osallistuvia olioita (ks. esim. Leino 1993: 80) tai nostavat näyttämölle täysin uusia olentoja, mikä heijastuu *ero*-ina argumenttirakenteessa.

Havainnollistan *ero*-pesyeessä ilmenevää ”epäsymmetriaa” kolmella esimerkkiparilla. Niistä ensimmäinen, esimerkit 2a–2c, ilmaisevat, miten kielenulkoisten olioiden koodaus vaihtelee perinteisin lauseenjäsennystermein, subjektin, objektin ja rektioadverbiaalin avulla tarkasteltuna. Alleviivattuna on olento, joka kokee muutoksen.

²⁷ Yksi tarkemmin luvussa 4 esittelemäni koodaussysteemin ongelmista on että järjestelmässäni on ”viralliset sarakkeet” vain kolmelle argumentille, X-, Y- ja Z-argumenteille. Kun *erottaa*-verbi (hyvin harvoin) esiintyy 4-paikkaisena, olen koodannut (tulostajaisen lausekkeen) lisäsarakeeseen Z², joka ei ole mukana analyysissä.

2a. Soini_{subjekti} **erotti** Pitkäsen_{objekti} puolueesta_{rekzioadverbiaali}.

2b. Pitkänen_{subjekti} **erosi** puolueesta_{rekzioadverbiaali} [koska Soini määräsi].

2c. Pitkäsen_{genetiivimääräite} **ero** puolueesta_{elatiivitäydennys} **on** positiivinen asia.

Esimerkkilauseet 2d–2e ilmaisevat 'entiteettejä erottavan tekijän' eli muuttujan VASTAUS syntaktista vaihtelua. Alleviivattuna on lauseke, joka spesifioi tämän muuttujan.

2d. Kissa ja koira **eroavat** siinä, että kissa on kissaeläin

2e. Kissaa ja koiraa **erottaa** se, että kissa on kissaeläin

2f. Kissalla ja koiralla **on vissi ero**. Kissa on kissaeläin

VASTAUS (ks. luku 4.4.3)

Ilmenee, että esimerkissä 2d muuttujan VASTAUS spesifioi adverbiaalitäydennys, 2e:ssä subjekti ja 2f:ssä itsenäinen lause. Viimeinen esimerkkipari, 2g–2h, ilmaisee syntaktis-semanttisen assosiatiivisuuden sijaan lekseemitasolta kumpuavaa lähisynonymiaa, nimittäin muuttujan RESIPROOKKI lekseemitason spesifikaation variaatiota.

2g. **Erotatko** "juottolan" ja "viihderavintolan" toisistaan?

2h. **Teetkö eron** "juottolan" ja "viihderavintolan" välillä?

Vastavuoroistava kielelementti eli RESIPROOKKI (ks. luku 4.2.2)

Ilmenee, että esimerkissä 2g sana *toisistaan*²⁸ esiintyy erosijaisen rektioadverbiaalin paikalla täydennyksenä, jonka läsnäolo ei ole kielipillisesti välttämätöntä. Esimerkissä 2h sana *välillä* taas on kielipillisesti välttämätön, sillä se esiintyy *ero*-nominin täydennyksenä toimivan lausekkeen "juottolan" ja "viihderavintolan" edussana.

Olen ratkaissut ongelman luopumalla täydennyksiä koodaavien muuttujien koodauksessa perinteisistä kielipin käsitteistä, nimittäin semanttisista rooleista (joita ovat muun muassa agentti, patientti ja lähde) mutta myös kielipillisistä rooleista eli rooleista, joilla on lauseenjäsennystehtävä (joita ovat muun muassa subjekti, objekti ja obliikvi) (ks. Leino 2005: Luento 4–6; Pajunen 2001: 83–85). Näiden sijaan hyödynnän Pajusen (2001) hengessä X-, Y-, ja Z-argumenteiksi nimeämiäni kategorioita²⁹. Argumentit esittelen tarkemmin luvussa 4.1. Ratkaisun seuraus on, että sivuutan tilastollisessa analyysissäni *ero*-lekseemien täydennysten syntaktisen korrelaation. Tämä tarkoittaa, että Y-argumentin spesifioi transi-

²⁸ Luvussa 2.2.2 totesin, että kaikki muuttujaluokkani manifestoivat prototyyppisiä kategorioita. Koodauksessa tämä tarkoittaa muun muassa, että olen määritellyt esimerkiksi muuttujan RESIPROOKKI mallijäseneksi *ero*-verbien täydennyksenä esiintyvän *toisistaan*-lekseemin. *ero*-nominissa esiintyvä *välillä*-lekseemi on hyväksytty luokkaan RESIPROOKKI *toisistaan*-sanaan liittyvän syntaktisen-semanttisen analogian nojalla.

²⁹ Pajunen (2001: 89) kuvaa käyttämiensä argumenttien syntaktis-funktionaalista vastaavuutta näin: "Normitapauksessa yksipaikkaisen verbin x-argumentti toimii subjektin funktiossa; kaksipaikkaisen verbin x- ja y-argumentit subjektin ja objektin tai rektio-obliikvi funktiossa; kolmipaikkaisen verbin z-argumentti on obliikvi (monissa kielissä se on objekti-2)." Sama ilmiö heijastuu *ero*-pesyeen argumenttien vaihtelussa (ks. luku 4.1).

tiivilauseessa objekti (2e), intransitiivilauseessa subjekti (2d) ja kopulalauseessa *ero*-nomin rinnasteinen täydennyslauseke (2f). *ero*-nominin täydennyksiksi tulkitsemiani jäseniä kutsun *ensimmäiseksi* ja *toiseksi täydennykseksi* (vaikka ne olisivatkin *määritteitä*, ks. s. 46 alav. 49). Tulkitsen argumentit semanttisen assosiatiivisuuden nojalla, lekseemien välisen oliotason analogian kautta, mikä tarkoittaa, että Y-argumentti ilmentää kaikissa *ero*-lekseemeissä samaa tai ainakin samankaltaista osallistujaa. Skeemojen EROTTAA.IRTI, EROTA.IRTI ja ERO.IRTI kohdalla Y-argumentti poimii siten tarkoitodellisuudesta sen entiteetin ’jota työpaikasta yms. eroaminen koskee’, Z-argumentti entiteetin ’jonka yhteydestä erotaan’. Koodausratkaisun taustalla on hypoteesi, että samankaltaisia olioita ja tilanteita esiintyy kaikkien *ero*-lekseemien merkitysrakenteessa – syntaktinen koodaus vain vaihtelee.

2.4 Käyttäytymisprofiilit

Pro gradu -tutkielmani monimuuttujaisen analyysin lähtökohtana toimivat käyttäytymisprofiilit (*Behavioral profiles, BP*), joiden käyttelyn olen omaksunut Griesin (2006, 2012) ja Divjakin ja Griesin (2006, 2008) sekä osaltaan Arppen (2008) töistä – ja etenkin Griesin, joka nimittää itseään kognitiiviseksi kielentutkijaksi. Vaikka työni painopiste on yksimuuttujaisessa analyysissä, esittelen käyttäytymisprofiileihin perustuvan lähestymistavan yksityiskohtaisesti, sillä olen koodannut aineiston juuri monimuuttujaisen analyysin tarpeita ajatellen. Luku valottaa myös merkityskäsitykseni kontekstuaalista ja kollikatiivista luonnetta.

Teoreettisesti käyttäytymisprofiilien³⁰ taustalla on korpuslingvistinen näkemys siitä, että kielellisen ilmauksen merkitys ja funktio (tutkimuksessani monimerkityksisten sanojen semanttiset suhteet) voidaan selvittää määrällisten menetelmien avulla, kontekstuaalisia yhdessäesiintymiä (*co-occurrence*) tutkien³¹ tarkasteltavan ilmiön jakaumasta, distribuutiosta (ks. Divjak & Gries 2009b: 277; Gries 2012: 58–60)³². Griessin (2012: 57; Gries ja Otani

³⁰ Ajatus käyttäytymisprofiileista periytyy Hanksilta (1996: 75), jonka mukaan verbien semantiikka määräytyy suoraan verbin komplementtirakenteesta (*complementation pattern*).

³¹ Näkemys siitä, että merkin kontekstuaalisen vaihtelun avulla voidaan vangita merkin funktio, merkitys ja tehtävä, ei ole uusi. Esimerkiksi Firth (1957: 11) kirjoittaa: “[...] you shall know a word by the company it keeps”. Bolinger (1968: 127) taas, että “a difference in syntactic form always spells a difference in meaning”. Yksityiskohtaisimmin asian on ilmaissut Harris (1970: 758): ”If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution.” Mainittakoon vielä lainaus Cruselta (1986: 1): “the semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts.”

³² Samankaltainen lähestymistapa on Jandalla ja Solovyellä (2009: 1–2), jotka käyttävät käsitettä konstruktio-naalinen profiili (*constructional profile*) kuvaamaan synonyymisten nominien leksikaalisia kollokaatioita (*lexical collocations*). Konstruktio-käsitteen he (mts. 8) määrittävät kytkien sen Langackerin (1987: 58) symboliseen yksikköön (*symbolic unit*) ja Goldbergin (1995) konstruktio-käsitteeseen: ”a conventionalized pairing

2010: 141–142) mukaan kaikki korpuspohjaiset rakennetta ja semantiikkaa tutkivat työt pohjaavat oikeastaan kahteen oletukseen: siihen, että i) kielellisen ilmauksen ominaisuudet voidaan paljastaa distribuution avulla sekä siihen, että ii) rakenteellisia eroja havainnoimalla voidaan paljastaa funktionaalisia eroja (”eri muoto, eri merkitys”-oletus).

BP-lähestymistavassa distribuutio vangitaan käyttäytymisprofiilien avulla. Lakoffin (1987) merkityskäsityksestä nouseva taustaoletus on, että monimerkityksisen sanan merkitykset (*senses*) muodostavat verkon, jossa samankaltaiset jäsenet kytkeytyvät toisiinsa osoittamalla samankaltaista morfologista, syntaktista, semanttista tai muuta vastaavaa lingvististä käyttäytymistä (Divjak & Gries 2009: 61). Gries ja Otani (2010: 142) uskovatkin, että käyttäytymisprofiileihin pohjautuva lähestymistapa integroi kielentutkimukseen vahvan kognitiivisen tutkimusasetteen.

Käytännössä ”Griesin metodi” (2006, 2012) rakentuu Atkinssin (1987: 24) ID-merkeille (*ID Tags*). ID-merkki – joka vastaa työssäni kategorisen muuttujan alaluokkaa³³ – voi mitata mitä tahansa lingvististä piirrettä, oli se sitten fonologinen, morfologinen, syntaktinen, semanttinen – tai ekstralingvistinen (ks. Gries 2012: 60). Monimuuttujaisessa analyysissä lähdän hypoteesista, että *ero*-lekseemien ”eri merkitykset” muistuttavat – tai eivät muistuta – toisiaan käyttäytymisprofiileiltaan, vektoreilta, jotka käsittävät valittujen muuttujien skeemakohtaiset prosenttijakaumat. BP-lähestymistavan etu on, että sen avulla päästään eroon nominaaliasteikolla toimivien kategoristen muuttujien rajoituksista. Koska ID-merkki on suhdeasteikollinen muuttuja, siitä voidaan laskea sellaisia tilastollisia määreitä kuin keskiarvo, keskihajonta tai korrelaatio.

Niin kuin Arppe (2008: 6 ja siellä mainitut lähteet) toteaa, tällaiseen näkemykseen perustuva ajattelu pohjaa Firthin (1957) muodostamaan traditioon, jossa merkitys ymmärretään kontekstuaalisesti vastakohtana (de)kompositionaalisuudelle. Käyttäytymisprofiilien avulla tarkastellaankin kollikaatioita. Hoeyn (2005: 43) mukaan kollikaatioissa³⁴ ei ole kyse vain siitä, että leksikaalisella yksiköllä on taipumus yhdessäesiintyä toisen leksikaalisen yksikön kanssa, vaan leksikaalinen yksikkö saattaa esiintyä myös tietyn grammaattiseen muodon kanssa. Työssäni kollikaatiolla viitataan sanojen, konstruktoiden ja rakenteiden yhdessä esiintymiseen. Hypoteesini on, että *ero*-pesyeessä ”eri merkitykset” palautuvat viime kädessä erilaisiin ko(n)tekstuaalisiin profiileihin (vrt. Kishner & Gibbs 1997: 27–28).

of form and meaning in a language”. Käsitteen *constructional profile* he (mts. 10) määrittelevät seuraavasti: “the relative frequency distribution of constructions that a given word appears in”.

³³ Vaikkapa muuttuja PÄÄLUOKKA sisältää kaksi ID-merkkiä: PÄÄLUOKKA_AKTIIVI ja PÄÄLUOKKA_PASSIIVI.

³⁴ Korpuksessa tutkittavien frekvenssien havainnoinnissa puhutaan kollikaation lisäksi kollikaatiosta ja kollikaatiosta, jotka ovat osittain päällekkäisiä termejä (käsitteiden määrittelystä ks. Gries 2010b: 275).

Griesin käyttäytymisprofiilien luominen sisältää neljä vaihetta (ks. Gries 2012: 59–66), jotka selostan tässä siten kuin itse niitä olen soveltanut. Ensiksi kokosin aineistoni HS.fi-korpuksesta luvussa 1.2 esitetyin periaattein. Toisessa vaiheessa loin muuttujiston, josta tässä työssä on mukana 20 muuttujaa. Tämän jälkeen siirsin korpusesimerkit (yhteensä 1666 tietuetta) koteksteineen taulukointiohjelmaan (*Open Office Calc*), jonne loin koodaustaulun, jossa korpusesimerkit sijaitsevat riveillä ja muuttujat sarakkeilla. Loin yhteensä kolme eri taulua, jokaiselle lekseemille omansa. *R*:ssä loin monimuuttujaista analyysia varten lisäksi neljännen taulun, joka yhdistää edelliset kolme taulua. Tämän jälkeen koodasin muuttujien arvot. Koodasin jokaisella esimerkille myös tiedon siitä, mitä skeemaa sen sisältämä *ero*-lekseemillinen lause manifestoi (ks. luku 3.3). Suppeaa otosta *erota*-lekseemin koodaustaulua havainnollistaa taulukko 1. Taulukko 2 esittää monimuuttujaisessa analyysivaiheessa (ks. luku 6.2) käytetyt muuttujat alakategorioineen (vrt. s. 161, alav. 162)

Taulukko 1. Todella suppea otos *erota*-verbin koodaustaulusta.

Esimerkki	LUKU	SKEEMA	Y_SEM	Y_MUOTO	RAKENNE	...
[..] <u>kädellisten ja lihansyöjien linjat erosivat</u> toisistaan 60 miljoonaa vuotta sitten.	K1	EROTA. ABSTRAKTI	ABSTRAKTI	NP	KONNEKTORI	...
80-luvulla <u>pa-sairaiden vanhemmista erosi</u> 80%	K2	EROTA. PUOLISO	JOUKKO	NP	Y_JAOLLI.	...
<u>Pariskunta oli eronnut</u> aiemmin, mutta [..]	K3	EROTA. PUOLISO	JOUKKO	NP	Y_JAOLLI.	...
[..] ja jos euro alkaa hajota , niin parempi [on] <u>erota</u> aiemmin kuin myöhemmin.	K4	EROTA.IRTI	NA	NOLLA	NA	...

Taulukko 2. Klusterianalyseissä käyttämäni ID-merkkien luettelo.

Tyyppi	Muuttuja	ID-merkki eli alakategoria
semanttinen	X_SEM, Y_SEM, Z_SEM	ELOT ON, KONKRETIA, INSTITUUTIO, JOUKKO, YKSIÖ, NA
morfologinen tai syntaktinen	AIKA	LÄSNÄ, NA
	SYV	LÄSNÄ, NA
	KVANTTORI	LÄSNÄ, NA
	VASTAUS	LÄSNÄ, NA
	MODUS1	VÄLTÄMÄTTÖMYYS, MAHDOLLISUUS, NA
	MODUS2	KYKY, NA
	X_MUOTO, Y_MUOTO, Z_MUOTO	NP, PRON, RAKENNE, PERS, NA, NOLLA,
	RAKENNE	KONNEKTORI, VAIN_Z, VAIN_Y, Y_JAOLLINEN, REKTIO, REKTIO2, NA
	RESIPROOKKI	LÄSNÄ, NA
	PÄÄLUOKKA	AKTIIVI, PASSIIVI
	KIELTO	NEUTRAALI, KIELTEINEN
	MODUS	INDIKATIIVI, MUU
	TEMPUS	PREESENS, PERFEKTI, MUU
	VERBI_RAKENNE	ERO.VERBI, VERBIKETJU, MUU, ON.VAIKEAA
	LUOKKA	INTRANSITIIVI, TRANSITIIVI, KOPULA

Griessin (ks. 2012: 60–62) BP-menetelmän kolmannessa vaiheessa absoluuttiset esiintymäfrekvenssit muunnetaan yhteisesiintymätaulukoksi (*co-occurrence table*) eli tauluksi, joka esittää kunkin muuttujan kaikkien alakategorioiden (ID-merkkien) prosentuaalisen jakauman³⁵ suhteessa tutkittaviin skeemoihin. Taulukko 3 havainnollistaa tällaista taulua.

Taulukko 3. Yhteisesiintymätaulu *erota*-skeemojen prosentiosuuksista³⁶.

Muuttuja	ID-merkki	EROTA.PUOLISO	EROTA.IRTI	EROTA.KIRKKO	EROTA.ERI
Y_SEM	ELOTON	0.00	0.00	0.00	0.75
	INSTITUUTIO	0.00	0.20	0.00	0.06
	JOUKKO	0.51	0.10	0.28	0.09
	YKSIÖ	0.18	0.45	0.38	0.03
	NA	0.31	0.26	0.34	0.06
PÄÄLUOKKA	AKTIIVI	0.96	0.96	0.95	1
	PASSIIVI	0.04	0.04	0.05	0
...

Taulukkoa 3 luetaan pystysuunnassa siten, että sarakkeisiin muodostuu luokkien skeema-kohtaisen jakauman perusteella yksilöllinen vektori, joka taulukossa 3 havainnollistaa harmaa laatikko. Tätä vektoria kutsutaan käyttäytymisprofiiliksi. Silmäilemällä havaitaan, että esimerkiksi skeeman EROTA.ERI esiintymistä 75 % osuu ID-merkkiin Y_SEM_ELOTON (joka mittaa intransitiivilauseessa agentin semanttista ontologiaa) ja että skeeman EROTA.ERI profiili poikkeaa muiden skeemojen profiilista.

Neljännessä vaiheessa data analysoidaan tilastollisin työkaluin, työssäni hierarkkisen klusterianalyysin avulla. Analyysissä tutkija voi valita, mitkä muuttujat otetaan mukaan ja mitkä jätetään analyysin ulkopuolelle. Esimerkiksi Deshors (2010) suorittaa väitöskirjassaan kolme analyysiä: yhden semanttisilla muuttujilla, toisen morfologisilla ja kolmannen kaikilla muuttujilla (vrt. Arppe 2008: 119). Klusterien pohjalta tutkija voi päätellä, miten merkitykset ryhmittyvät eli mitkä merkitykset ovat toistensa kaltaisia ja mitkä eivät.

Tarkastelen monimuuttujaisia menetelmiä ja niiden tuottamia tuloksia tarkemmin luvussa 6 ja siirryn seuraavaksi tutkimaan sanakirjoja ja *ero*-skeemojen koodausratkaisuja.

³⁵ Käyttäytymisprofiileihin pohjautuvan agglomeratiivisen klusterianalyysin heikkous on Arppen mukaan (2008: 151, 257) juuri se, että elementtien (skeemojen) vertailu pohjautuu luokkien (ID-merkkien) kokonaisprosenttiosuuksiin. Esimerkiksi skeemassa EROTTAA.IRTI (ks. luku 3) yksittäisen luokan (vaikkapa muuttujan Y_SEM luokan ELOTON) prosenttiosuus lasketaan 124 esimerkilauseen pohjalta. BP-menetelmällä ei voida tarkastella yksittäisten koodausesimerkkien vaikutusta ilmiöön.

³⁶ R-ohjelman skripteissä prosenttiosuuden esiintyvät kahdeksan desimaalin tarkkuudella. Lisäksi tässä havainnollistamisen vuoksi esillä oleva muuttujan Y_SEM luokkajako on eri kuin työssäni sovellettu, eivätkä arvot vastaa työni arvoja.

3 *ero*-pesyeen monimerkityksisyyden työhypoteesit

Tässä luvussa erittelen analyysien selittävät muuttujat eli *ero*-lekseemien merkitysvariantit – eli skeemat. Skeemoihin liittyvät luokitusratkaisut ovat työn laadullisin osa. Ei ole olemassa yksiselitteisiä kriteerejä sille, miten sanojen monimerkityksisyyttä tulisi jaotella. Yhdyn Kilgariffin (2003: 383) ajatukseen siitä, että ideaalitapauksessa lekseemin merkitysvaryuuden määrittäminen pohjautuisi korpukselta poimittuihin lause-esimerkkeihin, jotka ryhmiteltäisiin (tietyn muuttujiston nojalla) ”eri merkityksiksi” sen mukaan, ovatko analyysissä ilmenevät klusterit tarpeeksi isoja, erillisiä ja yhtenäisiä. Käytössäni ei kuitenkaan ole ollut sellaista tietokoneistettua laskentatetta, jolla olisin voinut esiryhmittää lause-esimerkkini valittujen kriteerien nojalla rykelmiksi – puhumattakaan, että olisin pystynyt suorittamaan klusterianalyysin siten, että selittävänä muuttujajoukkona toimisivat skeemojen sijaan aidot lause-esimerkit. Jatkotutkimusta silmällä pitäen skeemojen subjektiiviseen koodaamiseen olisi relevanttia etsiä objektiivisempia toteutustapoja (ks. luku 8).

Manuaalisesti koodatut ja subjektiivisesti määritellyt skeemaluokitukset toimivat määrällisen koeasetelman tarpeista kumpuavana välttämättömänä ratkaisuna, jonka paikkansa-pitävyyden arviointi on yksi työn taustatavoitteita. Edellä todetuista syistä johtuen ja koska tarkoitukseni on selvittää, miten eri *ero*-lekseemien käyttötavat suhtautuvat toisiinsa paitsi kunkin *ero*-lekseemin sisäisessä semantiikassa (luku 5), alustavasti myös *ero*-lekseemien välillä (luku 6), olen ottanut tendenssiksi jaotella merkitystä pikemminkin liian tarkasti kuin liian suppeasti. Tarkan jaottelun ongelmia tulee esille luvussa 5, jossa joudun kimputtamaan skeemoja yhteen khiin neliö -testauksen takia. Sivuan ongelmaa myös tämän luvun lopussa.

Aloitan tarkastelemalla sitä, miten sanakirjat pilkkovat *ero*-lekseemien merkitystä (luku 3.1). Luvussa 3.2 esitän työni varsinaiset tutkimuskohteet skeemat eli ”alamerkitykset”. Luvussa 3.3 käyn läpi kriteerit, joiden perusteella olen koodannut aineiston skeemoihin.

3.1 Sanakirjojen luokitukset

Merkityksen jaottelussa sanakirjat tarjoavat hyvän lähtökohdan, jota täydennän intuition ja kandidaatin tutkielmani havaintojen pohjalta. Tutkimani *ero*-lekseemit sisältävät tarkastelemieni sanakirjojen valossa useita eri merkityksiä. Koska tutkin nykykieltä, keskitän tarkasteluni tuoreimpiin sanakirjoihin. Niitä ovat Nykysuomen sanakirja (NS) ja Kielitoimiston sanakirja (KS) joka pohjautuu Suomen kielen perussanakirjaan (PS). Aloitan *ero*-nominista (luku 3.1.1), josta siirryn *erota*-verbin (luku 3.1.2) kautta *erottaa*-verbiin (luku 3.1.3).

3.1.1 ero-lekseemi sanakirjoissa

ero-nominin sanakirjamerkitykset ovat esillä taulukossa 4.

Taulukko 4. ero-verbin merkitykset sanakirjoissa.

Kielitoimiston sanakirja	Nykysuomen sanakirja
<p>1. eroaminen; erillään olo. <i>Ero virasta, virkaero Erik. Ottaa, saada ero</i> [=avioero] puolisostaan</p> <p>2. eroavuus, erilaisuus – <i>Tehdä ero joidenkin välillä</i> pitää erilaisina, erillään, erottaa toisistaan.</p> <p><i>leik. Eläköön se pieni</i> [miehen ja naisen] <i>ero!</i></p> <p><i>Erik. etäisyys, välimatka, lukumäärien t. mitattavien ominaisuuksien määrän erilaisuus t. erotus</i></p>	<p>1. eroaminen, eräminen; erillään olo, eronneena t. erotettuna oleminen.</p> <p>– Saada e., hakea e:a puolisostaan. Antaa e. miehelleen, vaimolleen. Ottaa, saada e., pyytää e:a virasta.</p> <p>2. eroavat ominaisuudet, eroav(ais)uus, erilaisuus – <i>tehdä ero joidenkin välillä</i> pitää eri asemassa, erillään, erottaa toisistaan</p> <p>3. etäisyys, välimatka; lukumäärien t. joidenkin mitattavien ominaisuuksien välimäärä, differenssi</p>

Sekä Kielitoimiston sanakirja että Nykysuomen sanakirja esittävät *ero*-sanalle lähes samat merkitykset, joita on laskutavasta riippuen 3–5. Yhteisiä merkityksiä ovat 1) 'eroaminen; erillään olo', 2) 'eroavaisuus erilaisuus' ja 3) 'etäisyys, välimatka, lukumäärien tai mitattavien ominaisuuksien määrän erilaisuus tai erotus'. Kohdan kaksi teksti-esimerkkien joukossa kumpikin esittää merkityksen 'tehdä ero joidenkin välillä, pitää erilaisina, erillään, erottaa toisistaan'. Merkitykselle 'ottaa, saada, pyytää eroa virasta' sanakirjat eivät esitä itsenäistä pykälää, vaan se on yhdistetty kohtaan yksi, 'eroaminen; erillään olo'³⁷.

3.1.2 erota-lekseemi sanakirjoissa

Sanakirjojen luokitukset, jotka ovat esillä taulukossa 5, muistuttavat toisiaan myös *erota*-lekseemin tapauksessa.

Taulukko 5. erota-verbin merkitykset sanakirjoissa.

Kielitoimiston sanakirja	Nykysuomen sanakirja
<p>1. mennä t. joutua erilleen, eroon jstak, erkaantua, irrota; etääntyä; lähteä, poiketa.</p> <p><i>Erik. Erotta elämästä</i> kuolla</p> <p>2. sanoutua irti työsuhteestaan, jnk yhteisön jäsenyydestä tms <i>Erik. ottaa avioero.</i></p> <p>3. olla erilainen, poiketa jstak.</p>	<p>1. Mennä t. joutua erilleen, eroon jstak (konkr. yhteydestä), irrota; kaikota, loitota, etääntyä; lähteä, poikeata – <i>Erik. kuolla</i></p> <p>2. Sanoutua irti jstak, katkaista välinsä, suhteensa jhk. E. virasta, koulusta. E. liitosta, puolueesta, kirkosta. E. miehestään, vanhemmistaan. Eronnut rouva.</p> <p>3. olla erilainen, poiketa jstak, pistää silmään; ks. myös eroava</p>

Seuraavat merkitykset ovat yhteisiä kummallekin sanakirjalle: 1) 'Mennä t. joutua erilleen,

³⁷ Leksikografien merkitysjaottelua on kritisoinut muun muassa Kilgarriff (2003). Sanakirjojen tekijät ovat Kilgarriffin (mts. 372) mukaan tietoisia siitä, että kahden "merkityksen" yhdistäminen (*lumping*) tai jakaminen (*splitting*) on lähtökohtaisesti subjektiivista toimintaa. Suuri osa merkityksen jaottelusta sijoittuukin todellisuudessa harmaalle alueelle "saman merkityksen" ja "eri merkityksen" välimaastoon. Leksikografit ovat hänen mukaansa hyvin kartalla muun muassa Lakoffin, Levinin ja Pustejovskin esiin nostamista ongelmista, jotka koskevat merkityksenjaottelua. Lisäksi sanakirjat tuotetaan sellaisessa toimituksellisessa hengessä ja sellaiselle yleisölle, joka suosii merkitystyyppien erittelyn perinnettä. Sanakirjojen odotetaan myyvän, eivätkä ne myy, elleivät ne sisällä odotuksenmukaisia listauksia sanojen eri merkityksistä ja käyttövaihtoehdoista. Kilgarriff mukaan nämä ovat syyt, joiden vuoksi leksikografit joutuvat kuvaamaan sanaston ikään kuin kaikkien sanojen semanttiseen rakenteeseen sisältyisiä joukko erillisiä, limittyttämiä merkityksiä.

eroon jstak, erkaantua, irrota; etääntyä; lähteä, poiketa', 2) 'sanoutua irti työsuhteestaan, jnk yhteisön jäsenyydestä tms.' ja 3) 'olla erilainen, poiketa jstak'. Näiden lisäksi kumpikin esittää merkityksen 'kuolla', jota ei esiinny aineistossani, eikä havaintojeni mukaan nykykielessä yleensä. Kielitoimiston sanakirjassa on erikseen mainittu merkitys 'ottaa avioero', joka Nykysuomen sanakirjassa on yhdistetty kohtaan kaksi, 'sanoutua irti jstak, katkaista välinsä, suhteensa jhk.'.

3.1.3 erottaa-verbi sanakirjoissa

Kielitoimiston sanakirja esittää *erottaa*-verbille laskutavasta riippuen 4–6 merkitystä ja Nykysuomen sanakirja 5–6. Sanakirjat erittelevät oikeastaan samat merkitykset mutta jaottelevat ja määrittelevät ne hieman toisistaan poikkeavasti. Merkitykset ovat esillä taulukossa 6.

Taulukko 6. *erottaa*-verbin merkitykset sanakirjoissa.

Kielitoimiston sanakirja	Nykysuomen sanakirja
1. saattaa erilleen, eroon, pitää erillään, erossa jostakin a. irrottaa, eristää, jakaa, loitontaa b. poistaa, sanoa irti; katkaista välit, estää läheisistä suhteista 2. pystyä havaitsemaan aistein, vars. saada näkyviin tai kuuluviin, nähdä, kuulla, haistaa, maistaa, tuntea; havaita eri yksilöiksi tms., erilaiseksi, tuntea 3. tehdä erilaiseksi, muista poikkeavaksi 4. ryhmittää, luokitkaa	1. saattaa erilleen, eroon, tai pitää erillään, erossa jostakin konkreettisesti yhteydestä, irrottaa, eristää, loitontaa 2. saattaa erilleen tai pitää erillään jostakin abstraktista yhteydestä; katkaista joidenkin suhteet, poistaa, sanoa irti, estää läheisistä suhteista 3. saada aisteillaan havaituksi, havaita, huomata, keksiä, vars. saada näkyviin tai kuuluviin, nähdä, kuulla 4a. tehdä jokin erilaiseksi kuin muut, muista poikkeavaksi 4b. Havaita eri olioiksi, erilaiseksi kuin muut, tuntea 5. jakaa t. yhdistää eri ryhmiin, jaottaa, ryhmittää, luokitkaa

Kummassakin sanakirjassa on merkitykset³⁸ 'irrottaa, eristää, jakaa, loitontaa', 'poistaa, sanoa irti; katkaista välit, estää läheisistä suhteista', 'pystyä havaitsemaan aistein, vars. saada näkyviin tai kuuluviin, nähdä', 'tehdä erilaiseksi, muista poikkeavaksi' sekä 'ryhmittää, luokitkaa'. Kielitoimiston sanakirjassa on erikseen merkitys 'havaita eri olioiksi, erilaiseksi kuin muut, tuntea', jonka Kielitoimiston sanakirja yhdistää merkitykseen 'pystyä havaitsemaan aistein jne.'. Kielitoimiston sanakirja esittää lisäksi merkityksen 1) 'saattaa erilleen, eroon, pitää erillään, erossa jostakin' jonkinlaiseksi ylämerkitykseksi, josta merkitykset 1a) 'irrottaa, eristää, jakaa, loitontaa' ja 1b) 'poistaa, sanoa irti; katkaista välit, estää läheisistä suhteista' kumpuavat. Nykysuomen sanakirja on ratkaissut saman luokittelukysymyksen antamalla abstraktille ja konkreettiselle toiminnalle omat pykälänsä.

³⁸ Parafrasit ovat Kielitoimiston sanakirjasta.

3.1.4 Yhteenveto

Tutkimieni sanakirjojen mukaan *ero-*, *erota-* ja *erottaa-*lekseemeillä on joitakin kaikille yhteisiä merkityksiä. Ne kaikki koodaavat merkityksiä, joiden skemaattisia yleistyksiä nimitän 'ominaisuuksien ja laadun eroavaisuuden skeemaksi', 'irtisanomisen skeemaksi', 'konkreettisen tai abstraktin eroamisen skeemaksi'. *erota-*verbi ja *ero-*nomini koodaavat kumpikin 'parisuhteen päättymisen skeemaa'. *erottaa-*verbi ja *ero-*nomini puolestaan koodaavat kumpikin erilaisia agentuaalisia kognitiivisia tiloja, joita nimitän 'kognitiivisen toiminnan skeemaksi'. Näitä ovat muun muassa erojen ja yhtäläisyyksien havaitseminen sekä luokittelu.

3.2 Taustatiedon pohjalta luodut skeemat

*ero-*lekseemien skeemojen muodostama kategoria toimii työssäni niin sanottuna riippuvana eli selitettävänä muuttujaryhmänä, joiden vaihtelua pyrin selvittämään riippumattomien eli selittävien muuttujien avulla (ks. esim. Nummenmaa 2004). Riippumattomia muuttujien esittely on luvussa (4). Aloitan tarkasteluni *ero-*lekseemistä (luku 3.2.1), josta etenen *erota-*lekseemin skeemojen (luku 3.2.2) esittelyn kautta *erottaa-*verbiin (luku 3.2.3). Tässä luvussa esitellyt skeemaluokitukset pohjaavat lukujen 3.1 sanakirja-artikkeleihin, kandidaatin tutkielmani tuloksiin (ks. luku 1.3) sekä kieli-intuitioon.

3.2.1 *ero-*nominin skeemat

*ero-*lekseemille määritin 8 skeemaa, joiden notaatiot ovat ERO.YMMÄRTÄÄ, ERO.HAVAITA, ERO.NÄHDÄ, ERO.IRTI, ERO.PUOLISO, ERO.ERI, ERO.TEHDÄ.ERO ja ERO.NUMEROT. Sanakirjojen pohjalta nousevia skeemoja ovat ERO.NUMEROT, ERO.IRTI, ERO.PUOLISO, ERO.ERI ja ERO.TEHDÄ.ERO. Näistä viimeisessä on kyse rakenteesta, jossa *ero-*lekseemi toimii *tehdä-*verbin objektina. Vaikkei kumpikaan tarkastelemistani sanakirjoista anna rakenteelle itsenäisen pykälän asemaa, molemmat nostavat sen esille. Se voi heijastaa sitä, että kyseessä on nykykieleen vakiintunut kollokaatio, jolla on konventionaalisen yksikön asema: kielenkäyttäjä ei tarvitse rakenteen hallitsemiseen konstruktivista ponnistelua, vaan rakenne on ikään kuin valmiina tuotepakkauksena kielellisten ilmausten varastossa (Leino 1993: 61).

Skeemat ERO.NÄHDÄ, ERO.YMMÄRTÄÄ ja ERO.HAVAITA syntyivät *ero-*nominin koodausvaiheen havainnoista, kun huomasin, että *ero-*nomini esiintyy usein sellaisten havaitsemista ja kognitiivista toimintaa kuvaavien verbien objektina kuin *nähdä*, *huomata*, *tietää*, *tajuta* ja *ymmärtää* (vrt. Pajunen 2001). Halusin selvittää, voisiko monimuuttujaisessa analyysissä edellä mainittujen ilmaisutyyppien ja eräitten *erottaa-*verbin ilmaisutyyppien välillä

ilmetä assosiaatiota. Lisäksi erotin mikroskeemat ERO.NÄKYÄ (n=4), ERO.KIRKKO (n=4) ja ERO.KONKRETIA (n=2), jotka on analyyseissä kuitenkin yhdistetty kategorioihin ERO.ERI ja jälkimmäiset skeemaan ERO.IRTI³⁹. *ero*-nominin skeemat ovat esillä taulukossa 7.

Taulukko 7. *ero*-nominin skeemat.

Skeema	Tekstiesimerkki	Σ
ERO.ERI	Tyttöjen ja poikien välillä voi olla suurikin ero yläkoulussa.	333
ERO.IRTI	Lapuanliike vaati koko hallituksen eroa .	36
ERO.PUOLISO	Vanhempieni ero on parasta.	37
ERO.NUMEROT	10 pisteen ero on pääsykokeessa valtava.	47
ERO.HAVAITA	Eivät huomaisi eroa , jos ei asiasta mainittaisi.	26
ERO.TEHDÄ.ERO	Pitää tehdä ero vapaaehtoisen ja pakotetun seksityön välille.	34
ERO.NÄHDÄ	Et näe eroa eläimen ja kehitysvammaisen ihmisen välillä?	20
ERO.YMMÄRTÄÄ	He tietävät, mikä on toden ja valheen ero .	61

3.2.1 *erota*-verbin skeemat

erota-verbille määrittelin 7 skeemaa. Suoraan sanakirjoihin pohjautuu neljä skeema, jotka ovat EROTA.ERI, EROTA.IRTI, EROTA.PUOLISO sekä EROTA.KONKRETIA. Koodausprosessissa aineistosta kumpusi lisäksi kolme skeemaa. Niistä EROTA.KIRKKO mainitaan kummassakin sanakirjassa, joskin ilman itsenäisen pykälän statusta. En halunnut yhdistää sitä skeemaan EROTA.IRTI, vaan annoin sille itsenäisen aseman tutkiakseni, poikkeavatko nämä ilmaisutyyppit toisistaan merkittävästi vai eivät. Lisäksi poimin mikroskeemat EROTA.IRTI2 ja EROTA.ABSTRAKTI, joilla skeemaa EROTA.KONKRETIA kanssa, on niin vähän esiintymiä, ettei tilastollinen analyysi ole relevanttia. Analyyseissä olenkin joko yhdistänyt ne muihin luokkiin tai sivuuttanut ne⁴⁰. Taulukko 8 esittää *erota*-verbin skeemat.

Taulukko 8. *erota*-verbin skeemat.

Skeema	Tekstiesimerkki	Σ
EROTA.ERI	Miten kohtuuton verottaminen eroaa varastamisesta?	206
EROTA.IRTI	Britannian puolustusministeri erosi kohuystävän takia.	155
EROTA.KIRKKO	Jos joku haluaa erota kirkosta niin, erotkoon .	116
EROTA.PUOLISO	Lesbot eroaa siinä kuin heterot, mutta homot ei.	73
EROTA.KONKRETIA	”Hei”, sanoiasentaja, kun erosimme .	2
EROTA.ABSTRAKTI	Apinalajien linjat erosivat toisistaan miljoonia vuosia sitten.	3
EROTA.IRTI2	Sitten kun kirkko eroaa valtiosta kokonaan.	8

3.2.1 *erottaa*-verbin skeemat

erottaa-verbille määritin 9 skeemaa, jotka ovat esillä alla olevassa taulukossa 9.

³⁹ Skeema ERO.IRTI sisältääkin aineksia, jotka puuttuvat skeemoista EROTA.IRTI ja EROTTAA.IRTI, samoin kuin skeema EROTA.IRTI sisältää aineksia, jotka puuttuvat skeemoista ERO.IRTI ja EROTTAA.IRTI. Saman lopputunnuksen omaavien skeemojen johdostenvälistä semanttista heterogeenisyyttä voidaan vähentää pilkkomalla skeemoja yhä pienemmiksi mikroskeemoiksi, jolloin tilastollinen testaus puolestaan käy mahdolltomaksi.

⁴⁰ Jo kandidaatintyössäni ilmeni, ettei sanomalehtiympäristössä *erota*-verbiä esiinny lauseen *Erosimme siellä ja täällä* (Bulgakov 2008: 230) kaltaisen lähtemisen tai erilleen joutumisen ilmoittamiseen (jotka ovat kummassakin sanakirjassa *erota*-verbin ”ensisijaisia” merkityksiä).

Taulukko 9. *erottaa*-verbin skeemat.

Skeema	Tekstiesimerkki	Σ
EROTTAA.IRTI	Eikö Hesari voisi jo erottaa tätä surkeaa kolumnistia.	124
EROTTAA.ERI	Työn erottaa harrastuksesta se, että siitä saa palkkaa.	34
EROTTAA.ABSTRAKTI	Pirkanmaa on aikoinaan erotettu Satakunnasta.	32
EROTTAA.KONKRETIA	Miten hampurilaisten jauheliha on erotettu luista?	41
EROTTAA.VÄLILLÄ	Se niemimaalta, jota erottaa muusta Euroopasta meri.	3
EROTTAA.LUOKITTELU	JP jos kuka, erottaa nämä käsitteet toisistaan.	60
EROTTAA.IRTI2	Vahinko kyllä, mutta kirkkoa ei ole erotettu valtiosta.	32
HAVAITA1	Hän ei ollut erottanut tummunutta seurantapantaa.	28
HAVAITA2	Jos ostaja ei osaa erottaa vanhaa kalaa tuoreesta [...].	155

Skeemoista 7 pohjautuu suoraan sanakirjoihin. Niitä ovat EROTTAA.IRTI, EROTTAA.ERI, EROTTAA.KONKRETIA, EROTTAA.VÄLISSÄ, EROTTAA.LUOKITTELU, EROTTAA.HAVAITA1 ja EROTTAA.HAVAITA2. Lisäksi määritin skeemat EROTTAA.IRTI2 ja EROTTAA.ABSTRAKTI. Skeeman EROTTAA.VÄLISSÄ olen analyseissä yhdistänyt isompiin ryhmiin. Hypoteesini ei nytkään ole, että kaikilla ”merkityksillä” olisi konventionaalistuneen yksikön asema leksikossa. Sen sijaan tarkoitukseni on testata, miten erilaiset suhteelliset vakiintuneet ilmaisutyytit suhteutuvat toisiinsa. Ilmaisutyyteistä kiistattomimpia lienevät sanakirjoista poimitut, kun taas osa on kokeiluluontoisia, kuten skeema EROTTAA.IRTI2⁴¹. *erottaa*-verbin skeemoilla on sen verran esiintymiä, että monimuuttujainen analyysi valitsemillani menetelmillä on relevanttia, joskin lähinnä suuntaa-antavaa (vrt. Gries 2006; Divjak & Gries 2006).

3.3 Korpusesimerkkien koodaus skeemoihin

Koodasin jokaisella HS.fi-korpuksesta poimitulle lauseelle tiedon siitä, mitä skeemaa ne edustavat. Tämä on osa annotointiprosessia, jonka erittely jatkuu luvussa 4. Koska tutkimuksessa ei ole löydetty vedenpitäviä kriteereitä sille, milloin ilmaisulle voidaan antaa alamerkityksen status (ks. luku 2.2.2), en minäkään pyri esittämään sellaisia kriteereitä. Olen koodannut esimerkit puhtaasti kieli-intuitioni nojalla. Kielentutkijan kieli-intuition ei tietenkään pitäisi olla pelkkää ”mutu-tuntumaa”. Tämän luvun tarkoitus onkin ilmaista ne (semi)intuitiiviset luokituskriteerit, joihin annotointi pohjaa.

Aloitan skeemoista, jotka kuvaavat konkreettista liikettä tai suhteen katkeamista (luku 3.3.1), minkä jälkeen käyn läpi skeemat, jotka ilmoittavat laadun tai ominaisuuden eroavaisuutta (luku 3.3.2). Luvussa 3.3.3 tarkastelen abstrakteja, kognitiivisia ja aistillisia tilanteita ilmoittavia skeemoja *ero*-verbien osalta ja luvussa 3.3.4 *ero*-nominin osalta.

⁴¹ Skeeman EROTTAA.IRTI2 voisi yhdistää joko skeemaan EROTTAA.IRTI tai EROTTAA.ABSTRAKTI. Näistä ensimmäinen on ’irtisanomisen skeeman’ toteuma, jälkimmäinen ’konkreettisen tai abstraktin eroamisen skeeman’ toteuma. Kyse lienee ”vedenjakajatapauksesta”, skeemojen välimaastoon sijoittuvasta ilmaisutyytistä, jolla molempien ominaisuuksia (Raukko 1994: 53). Tällaisiin luokitusongelmiin klusterianalyysi sopii hyvin.

3.3.1 Konkreettisen liikkeen ja suhteen katkaisemisen skeemat

Suhteen katkaisemisen skeemoja ovat skeemat, joiden lopputunnisteena on IRTI, IRTI2, KIRKKO tai PUOLISO. Skeemoihin ERO.IRTI (n=32), EROTA.IRTI (n=163) ja EROTTAA.IRTI (n=124) koodaamani esimerkit ilmoittavat aspektiltaan perfektivisen prosessin, jossa ihmis-tarkoitteisen olion tai olioryhmän työsuhde, jäsenyys tai vastaava irtisanotaan. Kognitiivisessa kielipissa (ks. Leino 1993: 74, 85) prosessit ovat *perfektiivisiä* tai *imperfektiivisiä* siitä riippuen, liittyykö verbin osoittamaan tilanteeseen ajassa tapahtuva muutos vai ei. Perfektiivinen prosessi vastaa Lyonsin (1977: 483) dynaamista tilannetta ja imperfektiivinen staattista tilannetta. Malliesimerkkejä IRTI-skeemoista ovat seuraavat:

3a. Kaikki neljä johtajaa **pitäisi** kyllä **erottaa** tämän sotkun seurauksena. (K36)

3b. Asuntonministeri Kiurun **tulisi erota**. (K1067)

3c. Yleensä tässä vaiheessa **aletaan vaatia** liikenneministerin **eroa**. (K1251)

Kaikille esimerkeille yhteistä on, että ne poimivat henkilön (3b, 3c) tai joukon (3a), jonka työsuhteen pitäisi päättyä. Muutos voi motivoitua ”eroajaan” nähden kausatiivisesti (3a), omatahtoisesti (3b) tai jäädä spesifioimatta (3c).

Skeemojen EROTA.IRTI ja EROTTAA.IRTI erikoistapaukseksi hahmotin skeemat EROTA.IRTI2 (n=8) ja EROTTAA.IRTI2 (n=32), joita havainnollistavat esimerkit 3d ja 3e:

3d. Sitten kun kirkko **eroaa** valtiosta kokonaan. (K634)

3e. Norjassa valtio ja kirkko **erotettiin** toisistaan ja sama pitää meilläkin tehdä. (K139)

Esimerkkien 3d ja 3e ilmoittaman toiminnan seurauksena maailmassa tapahtuu muutos, joka ilmenee ”institutionaalisten” suhteiden uudelleen organisoitumisena. Skeemat EROTA.KIRKKO (n=116) ja ERO.KIRKKO (n=4) sen sijaan hahmottavat perfektivisen prosessin, jossa ihminen (3f) tai joukko (3g) katkaisee siteensä jonkinlaiseen uskonnolliseen yhteisöön. Esimerkit 3f ja 3g havainnollistavat koodausratkaisuja:

3f. Jeesus **ei ole eronnut** kirkosta. (K878)

3g. Nimikristittyjen pitäisi **erota** kirkosta, ei uskovaisten. (K914)

Esimerkeille 3f ja 3g yhteistä on, että kontekstista käy ilmi, että suhteen katkaisu koskee nimenomaan uskonnollista yhteisöä (*kirkko*), ei työsuhdetta tai vastaavaa. Toiminnan seuraus on, että ihminen ei enää kuulu kirkkoon tai vastaavaan. Perfektiivisiä ovat myös skeemat EROTA.PUOLISO (n=73) ja ERO.PUOLISO (n=37), jotka ilmoittavat avioliiton, parisuhteen tai vastaavan katkeamista. Esimerkit 3h ja 3i demonstroivat koodausratkaisuja.

3h. Nainen **eroaa** heti, kun miehen tulotaso laskee. (K593)

3i. Omassa perheessäni ainakin **ero oli** lapsille helpotus ja palautti rauhan. (K1508)

Kuten esimerkeistä ilmenee, PUOLISO-skeemoille ominaista on, että kontekstissa esiintyy parisuhteeseen liittyvää sanastoa. Muutos, joka tapahtuu, ei ensisijaisesti ole konkreettinen, vaikka merkitystaustaan voi liittyä konkreettisia ominaispiirteitä. Varsinaisen konkreettisen muutoksen skeemoja ovat ERO.KONKRETIA (n=2), EROTA.KONKRETIA (n=2) ja EROTTAA.KONKRETIA (n=44). Skeemoihin koodatut lauseet hahmottavat fyysisessä ulottuvuudessa tapahtuvan perfektisen (tai joskus imperfektisen) prosessin, jossa eloton olio (kuten *kivi*) tai elollinen olio (kuten *ihminen*) irtoaa konkreettisesta yhteydestä (mahdollisesti jotain apuvälinettä käyttäen) ilman että kyseessä on pari- tai työsuhteen, jäsenyyden tai vastaavan katkaiseminen. Esimerkit 3j–3m havainnollistavat koodausratkaisuja.

3j. Ja muutto **erottaisi** meidät vanhemmistani. (K466)

3k. Miten esimerkiksi hampurilaisten jauheliha **on erotettu** luista? (K178)

3l. Tällaista ihmiskauppaa minä harjoitan, sanoi asentaja kun **erosimme**. (K917)

3m. **Ero** lapsista **ei ole** maailmanloppu. (K1460)

KONKRETIA-skeemoihin koodatut korpushavainnot ovat siinä mielessä heterogeenisiä, että kaikki lekseemit koodaavat elollisten olioiden toimintaa (3j, 3l ja 3m), mutta vain *erottaa*-verbi koodaa myös elottomiin olioihin kohdistuvaa manipulaatiota (3k). Ihmisiä koodaavat lauseet eivät lisäksi ole puhtaasti spatiaalisia, sillä niihin taustatietorakenteeseen voi sisältyä esimerkiksi sosiaalisia tai emotionaalisia piirteitä. Yhteistä niille on, että kielenulkoisessa todellisuudessa tapahtuu konkreettinen muutos.

Skeeman EROTTAA.KONKRETIA alatyypiksi hahmotin skeeman, jonka notaatio on EROTTAA.VÄLISSÄ (n=3). Esimerkki 3n havainnollistaa skeemaa:

3n. Uraania sisältävän malmin **erottaa** pohjavesistä alle ranteenpaksuinen kumimatto. (K259)

Esimerkki 3n poikkeaa vaikkapa esimerkistä 3f (ks. s. 35) siinä, ettei mitään konkreettista muutosta tapahdu. Lause ilmoittaa *kvasiresultatiivisen tilan*, jossa objektin sijavaihtelu ei manifestoi suomen kielen yleistä resultatiivisuus : irresultatiivisuus -oppositiota⁴².

3.3.2 Laadun ja määrän eroavaisuuden skeemat

Skeemoihin, joiden lopputunnus on ERI eli skeemoihin ERO.ERI (n=333), EROTA.ERI (n=206) ja EROTTAA.ERI (n=34) koodatut esimerkit ilmentävät imperfektiivistä tilaa, joka ilmoittaa kahden tai useamman entiteettien välisen laadun tai ominaisuuden eroavaisuutta. Malliesimerkkejä ovat seuraavat:

⁴² Esimerkin 3n eräs tulkinta on, että totaaliobjekti ilmoittaa tilan täydellisyyttä (vrt. Huomo 2006: 510, 2001). Kvasiresultatiivisuus on kiinnostava ilmiö, jonka kuitenkin rajaan työstä (vrt. Tola 2010). Ilmiötä ovat hämmentelleet muun muassa Denison (1957: 168), Itkonen (1976: 182–185) ja Leino (1991: 143, 166).

4a. Minut **erottaa** perus keskiluokkaisesta suomalaisista se, että iltakaljan sijaan nautin iltajointin. (K262)

4b. Miten kohtuuton verottaminen **eroaa** varastamisesta? (K951)

4c. **Ero** romanin ja maalaisen välillä **on** se, että maalainen on vähintäänkin perintömetsän omistaja ja tuo rahaa Helsinkiin ostamalla sieltä kalliin asunnon. (K1531)

Yhteistä näiden skeemojen toteutumille on, että niiden ensisijaisena luentana on entiteettien välinen vertailu, joka voi kohdistua mihin tahansa oliotyyppiin, tässä ihmisiin (4a, 4c) tai toimintaan (4b). Skeeman ERO.ERI ”erityistyyppinä” on skeema ERO.NUMEROT (n=47), jonka esimerkeissä ensisijaisena luentana on lukumäärään, etäisyyteen tai määrän välinen erotus. Malliesimerkkejä ovat seuraavat:

4d. 10 pisteen **ero on** oikeustieteellisen pääsykokeessa valtava. (K1384)

4e. Naisilla vastaava ero **on** 6,8 vuotta. (K1392)

Esimerkissä 4d on kyse piste-erosta ja esimerkissä 4e ikäerosta⁴³.

3.3.3 Abstraktit, kognitiiviset ja aistilliset skeemat

Skeemat joiden lopputunniste sisältää notaation HAVAITA eli skeemat EROTTAA.HAVAITA, EROTTAA.HAVAITA2 ja ERO.HAVAITA liittyvät tunnistamiseen, havainnointiin tai kategorisointiin. Skeema EROTTAA.HAVAITA (n=28) ilmoittaa tilanteen, jossa havainnointiin kykenevä olio poimii ulkotodellisuudesta aistihavainnon eli näkee, kuulee, haistaa, maistaa, koskettaa tai havaitsee jotakin. Skeeman toteutumisissa *erottaa*-verbi on yleensä mahdollista korvata aistihavaintoa koodaavalla agenttiivisellä havaintoverbillä (ks. Huumo 2006b: 79), sellaisella kuin *maistaa* tai *nähdä* ilman, että merkitys isommin muuttuu. Esimerkit 5a–b havainnollistavat koodausratkaisuja.

5a. Hän **ei ollut erottanut** naaraan kaulassa ollutta tummunutta seurantapantaa. (K300)

5b. Jos vain joka kymmenes ihminen **erottaa** karjun hajun, onko oikein kuohita kaikki karjuporsaat? (K330)

Esimerkissä 5a ensisijaisena luentana lieene ’havaitseminen’, joka tapahtuu näköaistin ja esimerkissä 5b hajun varassa. Kun patientin spesifioi lause, tulkinta ei aina ole puhtaasti

⁴³ Kun kontekstissa ei esiinny lukusanoja, ei aina ole selvää, tulisiko havainto koodata luokkaan ERO.ERI vai ERO.NUMEROT. Epäselvempiä skeemaan ERO.NUMEROT koodattuja esimerkkejä ovat seuraavat:

5A Joskus [...] hintataso Suomeen verrattuna oli [...] alhainen, nyt **ero** Helsingin ja Tallinnan välillä **on** aika pieni. (K1420)

5B. Ryhmien välille muodostui tilastollisesti merkittävä **ero** aivoalueiden toiminnassa [...]. (K1533)

Esimerkissä 5A vertailun fokuksena on edellisestä lauseesta mainittu maiden välinen hintataso eli hintaero. Esimerkissä 5B tilastollisesti merkittävä ero voi viitata tilastollisesti merkitsevään tutkimustuloksen lukuun tai tietoon siitä, että ryhmät ovat tilastollisesti erilaisia. Vertailuksi voidaan nostaa havainto skeemasta ERO.ERI:

5C. Tämä **on** toinen merkittävä **ero ryhmien välillä**, sillä opiskelu on usein opiskelijan oma tietoinen valinta. (K1532)

Tässä yhden erilaisuuden syyn (’toisen merkittävän eron’) spesifioi sana *tämä*, jonka korrelaattina on aikaisempi tekstikatkelma.

aistillinen, kuten esimerkeissä 5c–5d, jotka lukeutuvat osittain jo skeeman EROTTAA.HAVAITA2 piiriin, vaikka ovatkin koodattu skeemaan EROTTAA.HAVAITA.

5c. Jollei **osaa** ampua saati **erota**, mitä on ampumassa, jääköön kotiin ja jättäköön homman paremmalleen. (K237)

5d. Kuka **erottaa** ruoasta, onko se ollut oikeastaan koiranruokaa, joka on vain jalostettu kebablihaksi? (K308)

Esimerkissä 5c on kyse paitsi näköaistin varassa havaitsemisesta, 'näkemisestä', myös kyvystä osata tunnistaa eri eläinlajeja; 'tietämiseen' liittyvää tulkintaa voimistaa patienttina toimiva epäsuorakysymys *mitä on ampumassa*. Esimerkissä 5d on kyse joko yksin makuaistin tai yhdessä näköaistin varassa tapahtuvasta havaitsemisesta. Kyse ei kuitenkaan ole puhtaasta aistihavainnosta, sillä epäsuora kysymyslause luo tässäkin tunnistamisen tai tietämisen luentaa.

Varsinaista 'tunnistamista' edustaa skeema EROTTAA.HAVAITA2 (n=155). Skeemaan koodatut korpusesimerkit hahmottavat tilanteen, jossa havainnointiin kykenevä olio poimii tarkoitteen tai tarkoitejoukon jonkin toisen tarkoitteen tai tarkoitejoukon yhteydestä ilmoitukseen 'havaitsemisen' lisäksi tai sen sijaan 'tunnistamista' tai 'identifioimista'. Tarkastelen aluksi esimerkkejä 5e ja 5f, jotka ilmoittavat tunnistamista.

5e. Ja totta kai poliisin silmä **erottaa** vanhojen tuttujen autot. (K386)

5f. Kaupunkilaisen, mistä maasta tuleekin, **erottaa** todennäköisemmin siitä, että ei riehu nakkikopilla kännissä. (K171)

Esimerkki 5e ilmoittaa nähdäkseni proposition siitä, että (kaikkien autojen joukosta) poliisi (osaa) tunnistaa tai 'tietää' vanhojen tuttujen autot. Kyse on paitsi aistihavainnosta myös 'identifioimisesta', joka edellyttää monenlaisia kognitiivisia prosesseja: tietämistä, muistamista, kategorisoimista ja niin edelleen. Esimerkissä 5f on kyse tunnistamisesta ja havaitsemisesta. Aistielin, jonka varassa toiminta on, eli näkö- ja kuuloaisti, paljastuu elävisijaisen tukipronominin varaan rakentuvasta lauseesta: *siitä, että ei riehu nakkikopilla kännissä*. Hieman toisenlaista tapausta edustavat esimerkit 5g, 5h ja 5i.

5g. Jos metsästäjät **eivät erota** hirveä koirasta tai edes toisesta metsästäjästä, niin [...] (K312)

5h. Useat **eivät erottaneet** kakkosolutta kolmosesta. (K402)

5i. Yli puolet vastaajista oli sitä mieltä, että eniten tunnistamista vaikeuttavat saman valmistajan ulkonäöltään liian samankaltaiset pakkaukset. Vastaajia huolettavat omien virheiden sijaan enemmän se, kuinka iäkäs lääkkeen käyttäjä **erottaa** pakkaukset toisistaan? (K303)

Esimerkit 5g ja 5h ilmoittanevat edellisiä esimerkkejä selvemmin proposition siitä, että agenttitarkoite ei (osaa) tietää, sanoa, kategorisoida tai tunnistaa, kumpi eläimistä tai oluista

on kumpi. Esimerkissä 5i tapaan kontekstissa esiintyy usein tunnistamisen ja kategorisoinnin semanttiseen kenttään liittyvää sanastoa, tässä *tunnistaa*-verbiä⁴⁴. Esimerkeissä 5j ja 5k, joissa patientin spesifioi lause, tulkinta muuttuu ”mentaalisemmaksi”:

5j. Päättäjät **eivät osaa erottaa** sitä, mikä on tarpeellista siitä, mikä ei ole. (K362)

5k. Itsenäisyys on ollut järjestetyssä avioliitossa sodan kanssa niin kauan, että harva **erottaa**, etteivät ne oikeasti ole ihan niin naimisissa, miltä näyttää. (K405)

Esimerkin 5j ensisijainen luenta lienee, että päättäjät eivät tiedä tai ymmärrä, milloin on kyse yhdestä ilmiöstä (’sität mikä on tarpeellista’), milloin toisesta ilmiöstä (’sität mikä ei ole tarpeellista’). Esimerkki 5k on samankaltainen: *erottaa*-verbin parafrasiksi sopisi *tajuta*-, *tietää*- tai *ymmärtää*-verbi. Yhteistä kaikille EROTTAA.HAVAITA ja EROTTAA.HAVAITA2 skeemojen havainnoille on, että muutos, mikäli sellaista voidaan puhua, ei tapahdu ulkomaa-ilmassa, vaan agenttikonstituentin spesifioiman olion mielessä.

Skeeman EROTTAA.LUOKITTELU (n=60) esimerkeissä on kyse tilanteesta, jossa ihmis-tarkoitteinen olio älyn tai muun henkisen kyvyn avulla ”irrottaa” entiteetin tai entiteettejä toisen entiteetin tai entiteettien joukosta ilmoittaen kategorisoinnista tai asioiden jäsentämistä kuvaavan tilanteen. Koodausratkaisuja havainnollistavat seuraavat esimerkit:

5l. **Pitää erottaa** kaksi asiaa, eli onko jokin asia oikein ja onko se laillista. (K184)

5m. Aalto-opin professori Mika Ojakangas kuitenkin **erottaisi** äärioikeiston ja -vasemmiston sen mukaan, vastustaako ryhmä jotain kansanosaa kuten äärioikeisto vai talouteen liittyviä asioita kuten ääriivasemmisto. (K343)

Esimerkeissä on mitä ilmeisemmin kyse kategorisoinnista, joka koskee ’asioita’ (5l) tai ilmiöitä ’ääriivasemmisto’ ja ’äärioikeisto’ (5m). Ero skeemaan EROTTAA.HAVAITA2 on, että toiminta ei ensisijassa edellytä havaitsemista. Sen sijaan se edellyttää mentaalista toimintaa, kuten päättelyä tai tietämistä. Tunnistamisen ja kategorisoinnin ero ei aina ole selvä, ja myös LUOKITTELU-skeeman toteumat voivat saada modaalisia luentoja:

5n. Kunpa valtaosa ”maahanmuuttokriitikoista” vain **osaisikin erottaa** toisistaan islamin ja islamismin. (K393)

5o. Ehkä Halla-aho ja Stiller ei **kykene erottamaan** asioina tekijää ja syytä tekoon? (K501)

Esimerkeissä *osata*- (5n) tai *kyetä*-verbit (5o) tuovat kykyyn liittyvää modaalista luenta. Esimerkit kytkeytyvät siinä mielessä skeeman EROTTAA.HAVAITA2 piiriin, että siihen

⁴⁴ Olen minimoinut esimerkkien ympäriltä kontekstia. Koodauksessa hyödynsin tietenkin koko kontekstin.

koodatuissa lauseissa on usein kyse ihmisen kyvystä kategorisoida ilmiöitä, tosin nimenomaan ulkomaailman ilmiöitä⁴⁵.

Hieman toisenlaisen tilanteen hahmottavat skeemat, joiden lopputunnisteena on ABSTRAKTI. Niitä ovat EROTTAA.ABSTRAKTI (n=32) ja EROTA.ABSTRAKTI (n=3). Koodausratkaisuja havainnollistavat esimerkit 5r–5u:

5r. Sitten sähköön myynti ja jakelu **erotettiin** toisistaan. (K235)

5s. Pk-seudun **voisi** toki **erottaa** muusta Suomesta. (K129)

5t. Ystäväni joka on turkkilaisesta perheestä tuskaili vielä lukiossa muslimien omaa uskononopetusta, joka **oli erotettu** ev.lut ja ateistien kursseista. (K238)

5u. Kädellisten ja lihansyöjien linjat **erosivat** toisistaan 60 miljoonaa vuotta sitten. (K510)

Esimerkissä 5r on kyse kahden toimintamuodon, 'myyntitoiminnan' ja 'jakelutoiminnan', eriyttämisestä, jonka seuraus ehkä on, että toiminnot suoritetaan eri ihmisten toimesta eri lokaatioissa. Esimerkissä 5s on kyse manipulaatiosta, jonka seurauksena PK-seutu ei enää lukeutuisi Suomeen, vaan kenties muodostaisi oman valtionsa. Esimerkissä 5t ilmoitetaan, että muslimien uskonopetus on järjestetty evankelis-luterilaisesta opetuksesta erillään – mahdollisesti eri aikaan, eri tilassa ja eri opettajan toimesta. Yhteistä esimerkeille on, että tosimaailmassa tapahtuu muutos, joka ei ole pelkästään spatiaalinen, kuten skeemoissa joiden alkutunnus on KONKRETIA, eikä toisaalta agenttitarkoitteen havaintoa tai tietämistä predikoiva. Viimeisessä esimerkissä 5u, joka kuuluu skeemaan EROTA.ABSTRAKTI (n=3), tosimaailmassa ei kuitenkaan tapahdu muutosta, vaan kyse on niin sanotusta subjektiivisesta liikkeestä (ks. Huomo 2001)⁴⁶.

⁴⁵ Aina ei ole selvää, pitäisikö havainto koodata skeemaan EROTTAA.HAVAITA2 vai EROTTAA.LUOKITTELU. Epäselvempiä tapauksia edustavat esimerkit 5D ja 5E:

5D. Puhut [...] katupölystä, pakokaasuista ym. tekijöistä keuhkosairauksien aiheuttajana; ilmeisenä tarkoituksena helliä ajatusta, ettei tupakan aiheuttamia sairauksia **voitaisi erottaa** näistä tekijöistä. (K232)

5E. Mä en tajua miten voi tukea sellaista armeijaa joka tappaa siviileitä, joka **ei erota** lapsia terroristeista ja tappaa sekä kristittyjä että muslimeita. (K427)

5D on koodattu skeemaan EROTTAA.LUOKITTELU, koska kyse on kai ensisijaisesti älyllisestä toiminnasta, tosin muitakin luentoja voi olla. 5E kuuluu skeemaan EROTTAA.HAVAITA2, koska kyse on sellaisesta lapsien ja terroristien tappamisesta, joka edellyttää näiden tunnistamista ja kategorisoimista. Parafrasiksi voi sopia *eronominin sisältävä lause: Armeija ei tee eroa lapsien ja terroristien välillä*. 5E:llä on kuitenkin monia luentoja. Yksi luenta on, että sekä lapsia että terroristeja otetaan mukaan armeijan toimintaan. Perifeerisempi "irvileuan" luenta on, että armeija ei konkreettisesti siirrä lapsia pois terroristien luota.

⁴⁶ Epäselvempää tapausta edustaa esimerkki 6A:

6A. **Olisiko** seuraavalla ideana laittaa 5-vuotiaat tekemään teollisuuden helppoja hommia ja sitä kautta **erottaa** työläiset ja kuhnurit jo oikeassa iässä. [...] Paluu 1900-luvun alun säätyläiselämään näyttää johtajia kiinnostavan. (K241)

Esimerkillä 6A on useita luentoja. Olen koodannut sen skeemaan EROTTAA.ABSTRAKTI, vaikka skeemat EROTTAA.KONKRETIA tai EROTTAA.HAVAITA2 olisivat myös olleet mahdollisia ratkaisuja.

3.3.4 ero-nominin mentaaliset skeemat

Skeemaan ERO.YMMÄRTÄÄ (n=61) olen koodannut korpushavainnot, joiden finiittiverbinä (n=40) esiintyy jokin seuraavista (joskus apuverbin kera): *ymmärtää* (n=20), *tietää* (n=7), *tajuta* (n=4), *ajatella* (n=2), *tuntea* (n=2), *mieltää* (n=2), *unohtaa* (n=1), *tiedostaa* (n=1), *mennä* 'tietää' (n=1). Malliesimerkkejä ovat seuraavat:

6a. Suomalaiset **eivät ymmärrä** kansalaisuuden ja maassa pysyvästi asumisen **eroa**. (K1184)

6b. Sijoittaminen on eri juttu kuin pyramidihuijaus. Ellet näiden kahden **eroa tiedä**, suosittelen että sijoitat vain omalle pankkitilillesi. (K1243)

Esimerkissä 6a on kyse agenttitarkoitteen älyllisestä kyvystä kategorisoida tai tunnistaa erilleen ilmiöt "kansallisuus" ja "maassa pysyvästi asuminen". Toinen luenta on, että kyse ei ole niinkään kategorisoinnista vaan ilmiöiden välisen eroavaisuuden tajuamisesta, toisin sanoen siitä, että agenttitarkoite ei ymmärrä sitä syytä, miksi yksi ilmiö poikkeaa toisesta ilmiöstä. Esimerkin 6b luentamahdollisuudet ovat samankaltaiset.

Skeemaan ERO.YMMÄRTÄÄ olen lisäksi koodannut esimerkit (n=21), joissa kopulalauseessa esiintyvä *ero*-lauseke toimii sellaisten transitiviverbien objektina kuin *tietää* (n=9), *ymmärtää* (n=6), *tulla mieleen* (n=2), *unohtaa* (n=2), *erottaa* (n=1), *voida kuvitella* (n=1). Malliesimerkkejä korpuksesta ovat seuraavat yhdyslauseet:

6c. Jo muinainen lääketiede **tiesi**, että myrkyn ja lääkkeen **ero on** annostelussa. (K1473)

6d. **Onko** kirjoittaja **ymmärtänyt, mikä ero on** pedofiililla ja opettajalla, joka kiusaa ja lanistaa oppilastaan. (K1523)

Esimerkkien 6c ja 6d luenta on yhdyslause rakenteesta huolimatta hyvin samanlainen kuin esimerkeillä 6a ja 6b. Nytkin kysymys on hallitsevan lauseen agenttitarkoitteen kyvystä tai valmiuksista älyllisesti hallita kahden kategorian ero. Varsinaista havaitsemista ilmoittava skeema on ERO.HAVAITA (n=26), jonka toteutumisissa finiittiverbinä esiintyy sellaisia leksemejä kuin *huomata* (n=17), *havaita* (n=3), *löytää* (n=3), *jääda huomaamatta* (n=1), *maistaa* (n=1), *tunnistaa* (n=1). Malliesimerkkejä ovat seuraavat:

6e. Onneksi **eroa** sähkötupakan ja oikean välillä **ei huomaa** edes hajusta. (K1181)

6f. **Vaikeahan** sitä **eroa on maistaa**, kun molemmat on lantrattuja tärkkelysliemiä. (K1100)

6g. Jos et kerro, kukaan **ei huomaa eroa**, jos teet makaronilaatikon soijarouheesta. (K1169)

Esimerkeissä 6e–6g on kysymys nolla-agentin kyvystä (6e) tai mahdollisuuksista (6f) aistipohjaiseen havaintoon, joka 6e:ssä spesifioituu hajuaistimukseksi, 6f:ssä ja 6g:ssä makuhavainnoksi (ja 6g:ssä ehkä lisäksi haju- tai näköhavainnoksi). Luentamahdollisuudet ovat samankaltaiset kuin skeemalla ERO.YMMÄRTÄÄ: kyse voi olla entiteettien välisen eroavaisuuden eksplikoinnista tai havaintopohjaisesta tunnistamisesta. Koska skeeman

ERO.HAVAITA lause-esimerkeissä esiintyy monia verbejä, ovat luennatkin jokseenkin erilaisia. *löytää-* ja *havaita-*verbejä kuvaavat esimerkit 6h ja 6i:

6h. Tässä on kuitenkin kyseessä niin mitättömät lisäykset luonnon taustasäteilyyn, ettei maallikoiden laitteilla **ole** mitään **mahdollisuutta havaita eroa**. (K1165)

6i. Normitallaaja **ei** täällä maan itänurkilla **löydä** marketista sikajalosteista yleensä muuta **eroa** kuin sen, onko sille tuputettu rypsiöljyä vai ei. (K1173)

Esimerkin 6h ykkösluenta lienee, etteivät maallikoiden laitteet ole tarpeeksi herkkiä poimimaan eroa (ihmisen aisteihin rinnastuvien) sensoriensa avulla. 6in *löydä*-verbillä on useita luentoja, joihin havainnon lisäksi liittyy muuta mentaalista toimintaa. Skeema ERO.HAVAITA on *ero*-skeemoista ehkä semanttisesti heterogeenisin ja samalla eksperimentaalisin, seikka, jonka otan huomioon tutkimustuloksia arvioidessani.

Skeeman ERO.NÄHDÄ (n=20) havaintoja yhdistää, että lauseen finiittiverbinä esiintyy *nähdä-* (n=19) tai *kokea*-verbi (n=1). Koodausratkaisuja havainnollistavat esimerkit 6j–6l:

6j. Sinä **et näe** mitään **eroa** eläimen ja vaikkapa kehitysvammaisen ihmisen välillä? (K1178)

6k. **Etkö** oikeasti **näe** mitään **eroa**, sillä että ihmiselle suoritetaan sydänleikkaus lääketieteellisestä syystä ja sillä että tyydytetään vanhempien uskonnon vaatimuksia. (K1179)

6l. **En koe** siinä mitään **eroa**, nähdäänkö minut Alkossa tai vaikka vaatekaupassa. (K1192)

Esimerkeissä 6j ja 6k *nähdä*-verbin ensisijaisena luentana ei ilmeisesti ole aistillinen vaan mentaalinen merkitys – idea siitä, tekeekö agenttikonstituentin spesifioiva olio ajatusmaailmassaan eron kahden ilmiön välille. *kokea*-verbin semantiikka (6l) tuo tulkintaan ”ruumiillisuutta”, mutta muistuttaa tarpeeksi *nähdä*-verbin esittämiä tilanteita kuuluakseen kategoriaan. Neljästi *ero*-nomini esiintyy *näkyä*-verbin kanssa, kuten skeemaan ERO.NÄKYÄ koodatussa lauseessa *Jo 600–700-luvuilla alkoi näkyä eroa idän ja lännen kirkon välillä* (K1094). Tässä intransitiivinen *näkyä*-verbi predikoi selvästi erilaisen tilanteen kuin agenttiivinen *nähdä*-verbi muistuttaen eksistentiaalilauseetta.

Skeemaa ERO.NÄHDÄ muistuttava skeema on ERO.TEHDÄ.ERO (n=36). Siihen koodatut korpusesimerkit hahmottavat tilanteita, joissa agenttitarkoite älyn tai vastaavan henkisen avun avulla kategorisoi ilmiöitä erilleen. Finiittiverbinä esiintyy jokin seuraavista verbeistä tai rakenteista mahdollisesti modaalisen apuverbin kanssa: *tehdä ero[a]* (n=30), *hyväksyä ja tunnustaa ero* (n=1), *pitää ero* (n=1), *keksiä eroa* (n=1), *sanoa eroa* (n=1). Prototyyppisesti finiittiverbinä toimii *tehdä*-verbi, kuten esimerkeissä 6n ja 6o:

6n. **En tee** mitään **eroa** siviili- ja/tai varusmiespalveluksen suorittaneiden välillä. (K1086)

6o. Armeija **ei osaa tehdä eroa** simputuksen, koulutuksen, sotilaskurin välillä. (K1213)

Esimerkki 6n esittää proposition siitä, että agenttitarkoite (’minä’) ei kategorisoi tai

tyypitä ihmisiä asepalveluksen suoritustavan perusteella. Esimerkissä 6o on kyse siitä, että armeijalla ei ole kykyä kategorisoida tai tunnistaa erilaisia ilmiöitä. Joskus ”erottelun fokus” on tutkijan riemuksi ilmipantuna niin kuin esimerkissä *Ajattelussaan ja argumentoinnissaan kannattaa tehdä ero valtion ja yksilön välillä* (K1610), jossa inessiivisijainen adverbiaalilauseke ilmoittaa, että kyse on nimenomaan mentaalisesta suorituksesta⁴⁷.

Aineistossa esiintyy lisäksi neljä esimerkkiä, joissa finiittiverbinä on muu kuin *tehdä*. Näitä havainnollistavat esimerkit 6p–6r:

6p. Ainakin minä **keksin** parikin **eroa** aikuisille miehille ja teinitytölle. (K1114)

6q. Alkoholi [...] lamauttaa aivoissa niitä alueita jotka vastaavat sosiaalisesti välttämättömistä estoista, eli siitä että tunnereaktion ja käytöksen välillä **voidaan pitää ero**. (K1585)

6r. Eli lopetetaan sopimusten ja lakien vastainen toiminta [...] ja **hyväksytään ja tunnustetaan ero** vapaaehtoisen prostituution ja pienen marginaalin välillä. (K1616)

Esimerkkien 6p *keksii eroa*, 6q:n *voidaan pitää eroa* ja 6r:n *hyväksytään ja tunnustetaan ero* hahmottavat kukin verbisemantiikastaan kumpuavan tilanteen. Kuuluakseen kategoriaan tilanteet ovat kuitenkin nähdäkseni riittävän samankaltaisia *tehdä*-verbin sisältävien lauseiden kanssa. Vaikka tulkinta olisi turhan avomielinen, ratkaisu tuskin vääristää tuloksia, sillä lauseet edustavat vain 8.9 %:a skeeman ERO.TEHDÄ.ERO toteutumista.

Käsiteltyäni skeemat siirryn nyt tarkastelemaan niin sanottuja selittäviä muuttujia.

4 Muuttujat

Tutkimuksessani on mukana 19 muuttujaa, joista valtaosa on joko sellaisenaan tai modifioituna (esimerkiksi alakategorioita yksinkertaistamalla) poimittu aikaisemmista tutkimuksista. Päälähteenäni ovat toimineet Karlssonin yms. (1980), Arppen (2008) ja osittain Griessin (2012) sekä Divjakin ja Griessin yhteistyöt (2006, 2008, 2009, 2009b). Koska tutkin johdospesyyttä ja koska olen yrittänyt sovittaa muuttujat yhtä aikaa kolmen lekseemin kuvaus-tarpeisiin, osaa muuttujista ei määrittelemässäni muodossa esiinny muissa töissä.

Aloitan tarkastelemalla muuttujia, jotka mittaavat *ero*-lekseemien täydennysten semanttisia ja morfologisia ominaisuuksia (luku 4.1), minkä jälkeen esittelen muuttujia, jotka

⁴⁷ Poikkeusta edustavat rakenteet, jossa ei esiinny agenttikonstituenttia (n=2), kuten esimerkki (7B) *Mitä eroa se tekee, onko promilleraja 0,2 vai 0,5?* (K1152) tai (7C) *Ei tee eroa ihmisten välillä, onko jossain yhteisössä vai erakkona* (K1287). Esimerkin 7B minimipariksi voidaan – enemmän tai vähemmän onnistuneesti – formuloida lauseet *Ihmisiä ei erota se, ovatko ne jossain yhteisössä vai erakkona* ja *Ihmiset eivät eroa toisistaan siinä, ovatko jossain yhteisössä vai erakkona*. Esimerkkiin 7C istuu ehkä 7B:tä luontevammin *erota*-verbillinen minimipari: *Miten se eroaa, onko promilleraja 0,2 vai 0,5*. Olen koodannut molemmat skeemaan ERO.ERI.

mittaavat argumenttirakennetta (luku 4.2), verbien ominaisuuksia ja kieltoa (luvut 4.3) ja lopuksi adverbiaaleja ja niin sanottuja yleisiä muuttujia (luku 4.4). Koska lukujen 5 ja 6 analyysi pohjaa yksinomaan koodausratkaisuihin, käyn ratkaisujani läpi varsin seikkaperäisesti. Koodausratkaisujen erittely on toimintaa, jonka voi käsittää aineiston laadulliseksi analyysiksi. Esitän tässä vaiheessa myös tietoja muuttujien jakaumasta.

4.1 X-, Y- ja Z-argumentit

Muuttujia, jotka mittaavat *ero*-lekseemien täydennyksinä esiintyviä kielenaineiksia (NP-lausekkeita, itsenäisiä lauseita jne.), on yhteensä kuusi. Nominatarkoitteiden semanttista ontologiaa mittaavia muuttujia ovat X_SEM, Y_SEM ja Z_SEM. Rakenteellisia ominaisuuksia mittaavia muuttujat ovat X_MUOTO, Y_MUOTO ja Z_MUOTO. Näistä muuttujista muotoutuu kolme paria, jotka ovat X_SEM + MUOTO (X-argumentti), Y_SEM + MUOTO (Y-argumentti) ja Z_SEM + MUOTO (Z-argumentti). Kuten tässä luvussa osoitan, X-, Y- ja Z-argumentit ilmentävät oikeastaan rooliarkkityyppejä ja niiden vaihtelu *ero*-lekseemeissä heijastaa argumenttirakenteissa tapahtuvia roolimutoksia (ks. Langacker 1999).

Tämä luku on kaksiosainen. Osioissa 4.1.1 ja 4.1.2 tarkastelen, millaisia kielenelementtejä kunkin lekseemin yhteydessä on koodattu X-, Y-, Z-argumenteiksi. Osissa 4.1.3 ja 4.1.4 tarkastelen muuttujien luokkajakoa ja koodausperiaatteita.

4.1.1 *erota*- ja *erottaa*-verbien koodausratkaisut

X-, Y- ja Z-argumentit mittaavat eri lausetyypeissä osittain eri lauseenjäseniä. *erottaa*-verbin sisältävässä transitiivilauseessa X_SEM ja X_MUOTO osoittavat agenttia; Y_SEM ja Y_MUOTO patienttia; Z_SEM ja Z_MUOTO lähdettä tai Y-argumentin jälkimmäistä osaa. *erota*-verbin sisältävässä intransitiivilauseessa X_SEM ja X_MUOTO koodautuvat tyhjiksi (NA), kun taas Y_SEM ja Y_MUOTO mittavat agenttia ja Z_SEM ja Z_MUOTO lähdettä tai Y-argumentin jälkimmäistä osaa. Käytän argumenteista jatkossa lyhenteitä X_ARG., Y_ARG. ja Z_ARG. *erottaa*- ja *erota*-verbien koodausratkaisuja havainnollistavat esimerkit 7a–7c:

7a. Useat X_ARG. **eivät erottaneet** kakkosolutta Y_ARG. kolmosesta.Z_ARG (K402)

7b. Laitetaanko rajat kiinni ja **erotaan** Eu:sta.Z_ARG. (K1039)

7c. JP X_ARG., jos kuka, **erottaa** nämä käsitteet Y_ARG. toisistaan. (K348)

Esimerkeissä 7a X-argumentin spesifoi sana *useat*, Y-argumentin sana *kakkosolutta* ja Z-argumentin sana *kolmosesta*. Mikäli rakenteessa esiintyy vain erosijainen täydennys,

olen koodannut sen Z-argumentiksi, kuten esimerkissä 7b. Kun objektitäydennyksen spesifioi joukkotarkoitteinen sana (7c), Z-argumentti koodautuu aina NA:ksi (ks. s. 61).

Kaikki Z-argumenteiksi koodaamani kielenelementit eivät ole *erota-* tai *erottaa-*verbien rektioadverbiaaleja, vaan osa kuuluu syntaktisesti katsoen transitiivilauseen objektilausekkeeseen (*erottaa-*verbi) tai intransitiivilauseen subjektilausekkeeseen (*erota-*verbi). Tämä johtuu siitä, että *ero-*johdosten täydennyksissä esiintyy systemaattista vaihtelua, jota ilmentää minimipari *Kissa eroaa koirasta : kissa ja koira eroavat (toisistaan)* (ks. s. 66). Y-argumentteina esiintyvät kaksiosaiset lausekkeet (jotka sisältyvät muuttujan RAKENNE luokan KONNEKTORI piiriin) olen koodannut esimerkin 7d havainnollistamalla tavalla:

7d. **Oppii erottamaan** ystävät_{Y_ARG.} ja viholliset_{Z_ARG.} (K491)

Tällaisissa lauseissa Y-argumenteiksi olen koodannut *ja*-sanaa (tai vastaavaa konjunktiota, ks. s. 64) edeltävän jakson, tässä *ystävät*, ja Z-argumenteiksi *ja*-sananjälkeisen jakson, tässä *viholliset*. Joskus täydennyksenä oleva lauseke koostuu useammasta kuin kahdesta elementistä, kuten esimerkissä 7e:

7e. Voitko kirkkain silmäin väittää, että ihmisen, ihmisen identiteetin_{Y_ARG.} ja identiteetin olennaisen sisällön_{Z_ARG.} **voi erottaa** toisistaan. (K190)

Tällöin olen koodannut Y-argumentiksi kaikki *ja*-sanaa edeltävät osat, tässä tapauksessa jakson *ihmisen, ihmisen identiteetin*. Z-argumentin spesifioi esimerkissä 7e jakso *identiteetin olennaisen sisällön*. Sanojen ja lausekkeiden lisäksi täydennyksinä esiintyy lauseita, kuten esimerkeissä 7f–7h:

7f. Päättäjät_{X_ARG.} kun **eivät osaa erottaa** sitä, mikä on todella tarpeellista_{Y_ARG.} sitä, mikä ei ollenkaan ole._{Z_ARG.} (K362)

7g. Katsoja_{X_ARG.} **ei enää erota** räiskiikö päin tietokoneen tuottamaa kuvaa_{Y_ARG.} vai oikeaa._{Z_ARG.} (K372)

7h. **Pitää erottaa** kaksi asiaa, eli onko jokin asia oikein_{Y_ARG.} ja onko se laillista._{Z_ARG.} (K184)

Esimerkissä 7f sekä Y- että Z-argumentin spesifikaatioina toimii *se*-tukipronomini + lause -rakenteinen konstruktio. Esimerkissä 7g Y- ja Z-argumentin spesifioi epäsuora kysymyslause, joka koostuu *vai*-konjunktion yhteen kytkemistä rinnasteisista jaksoista. Nämä tapaukset olen hahmottanut seuraavasti: esimerkissä 7g Y-argumentin spesifioi lauseke *räiskiikö päin tietokoneen tuottamaa kuvaa* ja Z-argumentin *räiskiikö päin oikeaa kuvaa*. Vastaavasti esimerkissä 7h olen jäsentänyt Y-argumentiksi lauseen ”määritteekseen” saavan lausekkeen *kaksi asiaa, eli onko jokin asia oikein* ja Z-argumentiksi *kaksi asiaa, eli onko se laillista*. Mikäli rinnasteisia elementtejä ei esiinny, on paljas lause tai tukipronominillinen

konstruktio koodattu kokonaisuudessaan Y-argumentiksi. X-argumentiksi koodautuu transitiivilauseessa finiittiverbin ykkösargumentti, esimerkissä 7f *päättäjät* ja esimerkissä 7g *katsoja*. Agenteiksi työssä käsitän NP-muotoiset perussubjektit, verbiin inkorporoidut subjektit, nesessiivirakenteiden genetiivisubjektit, partitiivisubjektit sekä joissain tapauksissa paikallissijamuotoiset tekijälausekkeet (ks. VISK § 910, 919)⁴⁸. Esimerkit 7j ja 7k havainnollistavat genetiivissä ja paikallissijassa olevia X-argumentteja:

7j. Jokaisen puolueen_{X_ARG}. **tulee erottaa** jäsenet, jotka syyllistyvät rasismiin_{Y_ARG}. (K425)

7k. Sinulla_{X_ARG}. **näyttää olevan vaikeuksia erottaa** perusteltu kritiikki_{Y_ARG}. haukkumisesta_{Z_ARG}. (K319)

Paikallissijassa olevia (habitiivisia) adverbiaaleja olen koodannut X-argumentiksi varauksella, eikä niitä montaa olekaan. *erottaa*- ja *erota*-verbin argumenttien koodaukseen liittyvää selontekoa jatkan luvussa 4.1.3 ja siirryn nyt *ero*-nominin koodauksen käsittelyyn. Se poikkeaa jonkin verran *ero*-verbien koodauksesta, minkä takia käsitelen sitä erillään.

4.1.2 *ero*-nominin koodausratkaisut

ero-lekseemiä esiintyy niin kopulalauseissa, intransitiivilauseissa kuin transitiivilauseissa. Kun *ero*-lekseemi esiintyy kopulalauseessa (n=338) muuttujat Y_SEM ja Y_MUOTO mittaavat *ero*-nominin ensimmäistä täydennystä ja Z_SEM ja Z_MUOTO toista täydennystä⁴⁹. Muuttujat X_SEM ja X_MUOTO jäävät tyhjiksi. Esimerkki 8a ilmentää Y- ja Z-argumenttien koodausta:

8a. Todellisuuden_{Y_ARG}. ja puheiden_{Z_ARG}. **ero on** kuin itäblokkissa aikanaan. (K1579)

Tässä ”protoesimerkissä” Y-argumentin spesifioi sana *todellisuuden* ja Z-argumentin sana *puheiden*. *ero*-nominin ensimmäiseksi ja toiseksi täydennyksiksi kutsumani argumentit esiintyvät aineistossani kuitenkin monenlaisissa sijamuotojen yhdistelmissä. Abstrahoin yhteensä 24 ”rakennetyyppiä”, joista tässä käsitelen keskeisimmät⁵⁰. Päädyin yleistykseen, jonka mukaan (ilmipannut) Y- ja Z-argumentit esiintyvät kuudessa tyypissä: olosijassa (tyyppi 1), genetiivissä (tyyppi 2), olosijan ja tulosijan yhdistelmässä (tyyppi 3), elatiivissa (tyyppi 4) sekä lisäksi lauseessa (tyyppi 5) tai tyhjinä (tyyppi 6).

⁴⁸ Tukimuuttujan X_SIIA – joka mittaa subjektin tyyppiä – jakauma on: perussubjekti (n=226), genetiivisubjekti (n=24), inkorporoitu agentti (n=58), paikallissijassa oleva adverbiaali (n=11).

⁴⁹ *ero*-nominin Y- ja Z-argumentit – etenkin genetiivimuotoiset seuralaiset – ovat kuitenkin VISK:n hengessä (ks. esim. § 439) yleensä *määritteitä*, eivät *täydennyksiä*. Käyttämäni termit *ensimmäinen* ja *toinen täydennys* ovat *ero*-verbien analogiasta kumpuavia skemaattisia kattokäsitteitä. *ero*-nominiin kytkeytyviä seuralaisia olisikin ”turvallisempaa” nimittää *laajennuksiksi* (VISK:n kattokäsite täydennyksille ja määritteille).

⁵⁰ Rakenteiden runsauden syy ei juonnu pelkästään siitä, että *ero*-nomini esiintyy monenlaisissa lausetyypeissä (ks. VISK 891§, asetelma 140: Lausetyyppejä) ja siten kymmenien eri verbien täydennyksinä, sillä pelkästään kopulalauseissa esiintyy 18 erilaista rakennetyyppiä.

Tyypissä 1 Y- ja Z-argumenttien spesifikaatiot ovat adessiivissa (n=82), inessiivissä (n=81) tai genetiivin ja adessiivin yhdistelmässä (n=94). Näitä eri sijamuotovaihtoehtoja havainnollistavat esimerkit 8b, 8c ja 8d:

8b. Puuttumisella_{Y_ARG.} ja huomioimisella_{Z_ARG.} **on** vissi **ero**. (K1526)

8c. Eikö ihmisissä_{Y_ARG.} ja Pahaisissa Patrioteissa_{Z_ARG.} **ole** mitään **eroa**?⁵¹ (K1170)

8d. Kas kun seksin_{Y_ARG.} ja rakastelun välillä_{Z_ARG.} **on** iso **ero**.⁵² (K1541)

Tähän tyyppiin kuuluvat myös esimerkkien 8e ja 8f havainnot:

8e. Vaarallisten aineiden kuljetuksessa **ei tehdä**⁵³ **eroa** sillä, onko raketti ohjautuva eli ohjus_{Y_ARG.} vai ei_{Z_ARG.} (K1197)

8f. En_{X_ARG.} **ole huomannut** mitään **eroa** siinä, onko possu syönyt kasvisrehua_{Y_ARG.} vai ei_{Z_ARG.} (K1120)

Esimerkeissä 8e-8f Y- ja Z-argumentin spesifioivat *se*-tukipronomini + lause -rakenne. Edellisessä adverbialitäydennyksen tukisana on adessiivissä, jälkimmäisessä inessiivissä.

Tyypissä 2 pelkästään Y-argumentti (n=35) tai molemmat (n=54) ovat genetiivissä. Esimerkit 8g ja 8h havainnollistavat koodausratkaisuja (ks. myös 8a):

8g. Omien vanhempien_{Y_ARG.} **ero oli** helpotus. (K1509)

8h. Simputuksen_{Y_ARG.} ja "kovan koulutuksen"_{Z_ARG.} **ero on** mitätön. (K1557)

⁵¹ On kiinnostavaa pohtia tyyppiin 1 suhdetta eksistentiaali- ja omistuslauseisiin. VISK:n (§ 891) mukaan ”monikäyttöisillä lausetyypeillä on lauseenalkuinen nominatiivimuotoinen perussubjekti, kun taas erikoislausetyypeissä on ensimmäisenä jäsenenä paikallis-, partitiivi- tai genetiivisijainen lauseke ja niillä joko ei ole subjektia tai se on vähemmän subjektimainen kuin monikäyttöisten lausetyyppien subjekti. Erikoislausetyyppien analysoinnissa onkin olennaisempaa nähdä konstruktio usean ominaisuuden kimpuna kuin jäsentää niitä perinteisten lauseenjäsenten avulla.” VISK:n (§ 893–895) mukaan prototyyppisen e-lauseen ehdot ovat: 1) verbinä *olla*, 2) teemapaikalla paikanilmaus (jonka korvaa omistuslauseessa habitatiivinen lauseke), 3) jaollinen subjekti partitiivissa, 4) kieltolauseessa partitiivi, 5) verbi ei kongruoi ja 6) subjekti aiemmin mainitsematon. Kun tarkastellaan esimerkkejä 8A–C, havaitaan, että ainakin ehdot 1, 2, 3, 4, 6 täyttyvät.

8A. Eli käsitteillä ”romaanit” ja ”tietokirjat” **ei ole** mitään **eroa**. (K1084)

8B. Yhteisöllisyydellä ja anarkialla **on** vissi **ero**. (K1631)

8C. **On eroa** insinööriellä ja D-insinööriellä. (K1180)

e-lauseen sijaan nämä voidaan myös katsoa omistuslauseen laajentumiksi.

Tarkastellessani *ero*-nominin koodausesimerkkien ensimmäistä jäsentä kopulalauseissa testimuuttujan ERO.ALKU avulla havaitsin, että *eroa*-muodolla yleisimmät alkukonstituentit ovat olosijainen lauseke (53.4 %), kysymyssana (21.9 %), kieltosana (12.3 %), kun taas *ero*-muodolla yleisimmät ovat olosijainen lauseke (19.6 %), *ero*-nomini (18.9 %), kvanttori (14.3 %), *olla*-verbi (14.3 %) ja *se ~ siinä ~ tämä* -lauseke (12.1 %). Kummallakin yleisin on siis olosijainen konstituentti, mikä puoltaa erikoislausetyyppitulkintaa. Huomion arvoista on myös, että *ero*- ja *eroa*-muodot koodautuvat lauserakenteisiin hyvin erilaisesti. Yksi tutkimuksessani heikkous on, etten aina erottele *ero*- ja *eroa*-muotoja skeemakohtaisesti. Koska työni fokuksena ovat kuitenkin skeemojen väliset suhteet myös johdosten välillä, en pidä tätä isona puutteena.

⁵² Tyyppiin 1 rakenne Y-arg._{genetiivi} + Z-arg._{genetiivi} + *välillä* olisi voinut sijoittaa myös tyyppiin 2 alle, koska argumentit ovat genetiivissä. Olen kuitenkin luokittanut sen tyyppiin 1, sillä ilman *välillä*-sanaa (eli muuttujan RESIPROOKKI spesifikaatiota) lauseet olisivat epäkieliopillisia, kuten *kas kun seksin ja rakastelun on iso ero*. () Lause muuttuu kieliopilliseksi sanajärjestyksestä vaihtamalla: *kas kun seksin ja rakastelun ero on iso*.

⁵³ Verbit *huomata* ja *tehdä* sekä mentaaliverbit *ymmärtää* ja *tajuta* ovat sillä tavalla oppositiossa, että edelliset voivat saada tukipronomillisen adverbialitäydennyksen, jälkimmäiset myös lausetäydennyksen.

Esimerkissä 8g esiintyy pelkästään Y-argumentti, jonka spesifioi lauseke *omien vanhempieni*. Esimerkki 8h poikkeaa esimerkistä 8d (*seksin ja rakastelun välillä on ero*) siinä, että 8h on kielio pillinen ilman adessiivisijaista *välillä*-sanaa, kun taas 8d:ssä *välillä*-sana toimii lausekkeen edussanana.

Tyypissä 3 Y-argumentin spesifioima sana on adessiivissa (n=8), inessiivissä (n=14) tai genetiivissä (n=3), partitiivissa (n=1) tai sitä ei ole (n=30) samalla, kun Z-argumentti on illatiivissa. Esimerkit 8i–8k havainnollistavat koodausratkaisuja:

8i. Maahan tullessaan rikollisella_{Y_ARG} **on** yksi **ero** tavalliseen turistiin verrattuna_{Z_ARG}: ajatukset. (K1530)

8j. Ihmisen summittaisessa tappamisessa_{Y_ARG} autolla **en näe** mitään **eroa** ihmisen summittamiseen tappamiseen_{Z_ARG} ampuma-aseella. (K1158)

8k. **Onko** työssäolevien suomalaisten_{Y_ARG} ainoa **ero** työttömiin_{Z_ARG} (ja heidän motivaationsa olla töissä), se että heillä on korkeampi etiikka. (K1595)

Esimerkissä 8i on kyse yhdistelmästä Y_arg.adessiivi (*rikollisella*) + X_arg.illatiivi (*tavalliseen turistiin verrattuna*), 8j:ssä yhdistelmästä Y_arg.inessiivi (*tappamisessa*) + Y_arg.illatiivi (*tappamiseen*) ja 8k:ssa yhdistelmästä Y_arg.genetiivi (*suomalaisten*) + Z_arg.illatiivi (*työttömiin*). Esimerkin 8j *autolla* ja *ampuma-aseella* ilmoittavat tekemisen tapaa. Yleisin Y-argumentin sija on kuitenkin tyhjä, kuten esimerkissä *Stalin suunnitteli juutalaisvainoja, joten ero natsimeininkiin*_{Z_ARG} *ei siinäkään suhteessa niin suuri ollut* (K1358).

Tyypissä 4 Z-argumentti on elatiivissa, kuten esimerkeissä 8l ja 8m:

8l. **On** intti+reservistä_{Z_ARG} **ero** kevein vaihtoehto hoitaa tuo typerys alta pois. (K1529)

8m. Kuka **on ollut vaatimassa** kaikkien_{Y_ARG} **eroa** kirkosta_{Z_ARG}? (K1241)

Y-argumentti, mikäli esiintyy rakenteessa, on genetiivissä (n=3), kuten esimerkissä 8m, tai nominatiivissa (n=4). Oma tyyppiinsä ovat lisäksi tapaukset, joissa Y- (ja Z-) argumentin spesifioi lause. Tätä tyyppiä 5 havainnollistavat esimerkit 8n ja 8o:

8n. **Et**_{X_ARG} **näe** mitään **eroa** aseen käytöstä_{FOKUS} jos ampuu oikealla aseella_{Y_ARG} ja videopelissä aseilla_{Z_ARG}? (K1195)

8o. Kukaan_{X_ARG} **ei huomaa eroa**, jos teet makaronilaatikon soijarouheesta_{Y_ARG}. (K1169)

Esimerkissä 8n Y- ja Z-argumenteiksi on koodattu alisteisen *jos*-lauseen *ja*-konjunktioilla yhteen kytketyt rinnakkaiset elementit *jos ampuu oikealla aseella* ja [*jos ampuu*] *videopelissä aseilla*. Elatiivisijainen lauseke *aseen käytöstä* on koodattu tukimuuttujaan FOKUS. X-argumentiksi koodautuu kieltoverbiin inkorporoitu agentti (*et*). On yleistä, että

toinen argumentista jää ilmi koodaamatta, mutta sisältyy käsitteistykseen⁵⁴, kuten esimerkin 8o kaltaisissa (yhdyks)lauseissa, joissa Y-argumentti saa arvon, mutta Z-argumentti koodautuu NA:ksi.

Kuudes tyyppi on niin sanottu nollakonstruktio, jota kuvastaa esimerkit 8p ja 8q:

8p. Oleskeluluvan saatuaan **tulee ero**, ja puoliset jäävät veronmaksajien vastuulle. (K1657)

8q. Semmonen pieni **ero** vaan **on**, että oikea lääkäri on pikkuisen opiskellutkin. (K1313)

Nollakonstruktiolla ominaista on, että sekä Y- että Z-argumentti puuttuvat lauserakenteesta (eli koodautuvat tyhjiksi, NA), kuten esimerkeissä 8p ja 8q on asian laita.

Kun *ero*-nomini esiintyy transitiivilauseessa (n=204) finiittiverbin kakkostäydennysnä eli objektina (n=162), muuttujat X_SEM ja X_MUOTO mittavat finiittiverbin ykköstäydennystä eli agenttia (kuten esimerkeissä 8o ja 8p). Muuttujat Y_SEM ja Y_MUOTO osoittavat *ero*-nominin ensimmäistä täydennystä ja Z_SEM ja Z_MUOTO toista täydennystä, kuten esimerkissä 8q:

8q. Minä_{X_ARG.} **en näe eroa** alkoholin_{Y_ARG.} ja muiden huumeiden_{Z_ARG.} välillä. (K1286)

Tässä X-argumentin spesifioi sana *minä*, Y-argumentin sana *alkoholin* ja Y-argumentin lauseke *muiden huumeiden*. En kuitenkaan aina ole automaattisesti koodannut *ero*-lausekkeen objektikseen saavan transitiiviverbin agenttia X-argumentiksi, sillä lähtökohtana *ero*-nominillisten transitiivilauseiden koodauksessa on *erottaa*-verbin malli ja siihen liittyvä transitiivilauseen konsepti. Sen pohjana voidaan nähdä arkinen biljardipallomalli, jossa lyöjä siirtää kepillä energiaa yhteen palloon, joka kimpoaa toiseen palloon, toinen kolmanteen jne. (Leino 1993: 113; Onikki 1994: 123). Biljardipallomallin konseptiin läheisesti liittyvä käsite on tekoketju (*action chain*), jossa väylämalli (lähde-väylä-kohde -skeema) yhdistyy Talmyn (1985) voimadynaamiseen malliin: tekoketjussa energianlähteenä toimiva agenttitarkoite (X-argumentti) kohdistaa energiaa patienttiin (Y-argumentti) ja aiheuttaa tämän tilassa muutoksen (Langacker 1991: 282–284; Leino mts. 113–114, 204 alav. 4; Onikki 1994: 123–124; Jääskeläinen 2004: 104 alav. 3). Tekoketjussa ensimmäinen jäsen siirtää energiaa toiselle jäsenelle, joka välittää sen kolmannelle jne⁵⁵. Intransitiivisessa *erota*-verbissä tekoketjun ensimmäinen jäsen, agentti, toimii paitsi voimanlähteenä myös voimankohdeena, ja ”energia-

⁵⁴ Tulkintani mukaan esimerkin 8o yhdyslauseen *Kukaan ei huomaa eroa, jos teet makaronilaatikon soijaruouheesta* sisältää (elliptoidun) ”piioelementin”, jota kuvastaa propositio ”kuin jos tekisit makaronilaatikon jauhelihasta”. Yleensä ilmipanematon elementti mainitaan kontekstissa ennen *ero*-lausetta.

⁵⁵ Kausatiivisten lauseiden voimadynaamiset mallinnukset, jotka esittävät arkikäsitteistykseen perustuvan konkreettisen manipuloinnin energian siirtona, soveltuvat erityisesti spatiaalisten ilmauksiin. Abstraktimpien suhteiden kuvaus energian siirtona on ongelmallisempaa, mutta se voidaan palauttaa konkreettisen prototyypin fyysikaalisten suhteiden laajentamiseen ja metaforisoimiseen. (Jääskeläinen 2004: 104 alav. 3.)

virta”, mikäli sellaisesta tässä yhteydessä edes voi puhua, kohdistuu agenttiin itseensä. Transitiivilause tuo tekoketjuun lisätekijän, agentin (X-argumentti), jolloin verbin valenssiluku kasvaa yhdellä (ks. VISK § 314).

Voimadynaamisen ketjun alkuunpanija – X-argumentti – voi kuitenkin kohdistaa energiaa patienttiin – Y-argumenttiin – hyvin erilaisin tavoin, eivätkä kaikki tavat muistuta tapaa, jolla *erottaa*-verbin agentti kohdistaa energiaa patienttiin, olkoonkin, että näitä tapoja on useita, sillä *erottaa*-verbi on voimakkaasti polyseeminen. Esimerkiksi skeeman EROTTAA.IRTI toteutumisissa agentti (milloin spesifioitu) edustaa oliota, jolla kielenulkoisessa todellisuudessa on valta irtisanoa patienttitarkoite. Lauseessa *Savonian hallitus päätti erottaa opiskelijan vuodeksi opinnoistaan* (K429) tällainen on *Savonian hallitus*. Esimerkit 8r–8s havainnollistavat tapauksia, joissa X-argumentti jää koodaamatta:

8r. *Lapuan liike*_{EI KODATTU} **vaatii** *koko hallituksen*_{Y_ARG.} **eroa**. (K1253)

8s. *Poliisin tietoon tulleiden tapausten epäsuhta*_{EI KODATTU} **ei nosta eroa** [...] tilastollisesti merkittäväksi. (K1201).

Lauseen 8r *vaatia*-verbin subjektilauseke *Lapuan liike* on jäänyt koodaamatta X-argumentiksi siksi, että *vaatia*-verbin ilmoittama poliittisen muutoksen vaade on vahvimmillaan kin painostamista (tai ehkä uhkailua), ei konkreettista irtisanomista (joskin homma voi kumuloitua tilanteeseen, jossa pääministeri todella *erottaa koko hallituksen*). Lauseessa 8s agenttilauseke *poliisin tietoon tulleiden tapausten epäsuhta* on jäänyt koodaamatta, koska se *ei nosta*-verbin ykkösargumenttina edusta mitään sellaista ominaisuutta, jota muuttujistoni mittaa. Koodaamatta olen lisäksi jättänyt kopulalauseiden predikaatiivit silloin, kun ne eivät täytä muuttujan VASTAUS kriteereitä (ks. luku 4.4.3), kuten esimerkissä *Silloin ero on paras ratkaisu*_{PREDIKATIIVI} (K1642) tai esimerkin *Hyvän ja pahan välinen ero on nykyisessä yhteiskunnassa hämärtynyt*_{PREDIKATIIVI} (K1426). Seuraavat esimerkit havainnollistavat tapauksia, joissa subjekti on poikkeuksellisesti koodattu Y-argumentiksi:

8t. Vaikka *vaimo*_{Y_ARG.} **vaatisi eroa**. (K1235)

8u. Suomalaiset eivät hanki lapsia, jos *he*_{Y_ARG.} elävät huonossa parisuhteessa tai **pohtivat eroa**. (K1293)

8v. Lautamiehen paikka avautui, kun *Roope Tammi*_{Y_ARG.} **pyysi siitä**_{Z_ARG.} **eroa** perhesyiden ja työkiireiden vuoksi. (K1257)

8w. *Useat kunnat*_{Y_ARG.} **suunnittelevat eroa** *kuntaliitosta*_{Z_ARG.} (K1210)

Esimerkin 8t *vaimo* ja 8u:n *he* (*suomalaiset*) olen koodannut luokkaan Y-argumentti, sillä *vaatia* ja *pohtia*-verbien subjekteina ne hahmottuvat entiteeteiksi, jotka kokevat parisuhteen päättymisen. Samalla ne assosioituvat *erota*-verbin Y-argumentteihin lauseissa,

jotka predikoivat halua tai tahtoa, kuten esimerkissä *Nainen oli ilmoittanut miehelle haluvansa erota* (K990) tai nesessiivisyyttä, kuten esimerkissä *Lopulta oli pakko erota, en halunnut juonitella lasten miehen kanssa* (K549). Esimerkin 8v olen koodannut Y-argumentiksi samasta syystä: *Roope Tammi* erosi lautamiehen paikaltaan (mahdollisesti siten, että joku muodollisesti erotti hänet). *ero*-lekseemin käyttö assosioituu täten *erota*-verbiin, ei *erottaa*-verbiin, mitä demonstroi minipari *Roope Tammi pyysi ~ päätti erota siitä : Roope Tammi pyysi ~ päätti erottaa itsensä siitä*. Viimein 8w:ssä *useat kunnat* edustavat *erota*-verbille tyypillistä, voimanlähteenä ja voimankohteena toimivaa Y-argumenttitarkoitetta, joka voi *itse erota kuntaliitosta* niin halutessaan.

Kun *ero*-nomini esiintyy alisteisessa lauseessa, joka toimii transitiiviverbin lauseobjektina (n=37), olen pääsääntöisesti koodannut X-argumentiksi hallitsevan lauseen subjektin. Y- ja Z-argumenteiksi olen koodannut alisteisen (kysymys)lauseen (rinnasteiset jaksot). Tämä koodausratkaisu poikkeaa yleisestä linjastani, sillä tarkastelen yhdyslauseen ominaisuuksia. Mentaaliset ja loogisia asiayhteyksiä ilmaisevat verbit (kuten *ymmärtää* tai *tietää*) ja niiden lausetäydennykset muodostavat kuitenkin kokonaisuuden, jonka osien tarkastelu toisistaan erillään ei ole semanttisesti mielekästä. VISK:n (§ 472) mukaan tällaisissa tapauksissa alisteinen kysymyslause esittää proposition, jonka sisältö on joltakin osin avoin, ratkaisematon. Lause, jonka täydennyksenä alisteinen kysymyslause on, ilmaisee tämän ratkaisemattoman asian käsittelyä. Yhdyslauseen osien kytköksisyyttä heijastelee sekin seikka, että yhdyslause, joka sisältää alisteisen kysymyslauseen, tulkitaan funktioltaan ensisijaisesti päälauseen mukaan siten, että alisteisen kysymyslauseen sisältävä väitelause ei yleensä ole kysyvä (ks. VISK § 1703–1704). Aineistossani verbeillä, jotka voivat saada interrogatiivimuotoinen lausetäydennyksen (8x, 8z), esiintyy myös *ero*-lausekkeen varaan rakentuvia täydennyksiä, joita havainnollistavat esimerkit 8y ja 8å.

8x. Kun minä katson elokuvaa, niin kyllä minä X_ARG. **tiedän, mikä on** elokuvan Y_ARG. ja dokumenttielokuvan Z_ARG. **ero**. (K1406)

8y. Normitallaaja X_ARG. **ei** edes **tiedä** puoliperän Y_ARG. ja täysperän Z_ARG. **eroa**. (K1277)

8z. **Onko** kirjoittaja X_ARG. **ymmärtänyt, mikä ero on** pedofiililla Y_ARG. ja opettajalla joka kiusaa ja lannistaa oppilastaan? Z_ARG. (K1523)

8å. Eivät suomalaiset ole idiotteja, jotka X_ARG. **eivät ymmärrä** oikean Y_ARG. ja väärän Z_ARG. **eroa**. (K1193)

Esimerkkien 8x ja 8y *ero* ei ole niin suuri kuin voisi ajatella: kummassakin on kyse *ero*-nominin ensimmäisistä (*elokuvan, puoliperän*) ja toisista (*dokumenttielokuvan, täysperän*) täydennyksistä, samoin esimerkeissä 8z ja 8å. Lauseenjäsenyksellisesti *ero*-lausekkeet

ovat objekteja. 8y:ssa ja 8å:ssa *ero*-lauseke esiintyy ”paljaana”, kun taas 8x:ssä ja 8z:ssa *ero*-lauseke kytkeytyy kopulaverbiin ja kysymyssanaan.

Kun *ero*-lauseke esiintyy transitiivilauseen (n=5) tai intransitiivilauseen (n=52) subjektina, X_SEM ja X_MUOTO koodautuvat NA:ksi, kuten lauseessa *Ero*_{subjekti} *auttaisi myös muita euromaita*_{objekti}, jossa myös Y- ja Z-argumentit koodautuvat NA:ksi, sillä objektikonstituenttia jää koodaamatta. Koodaamatta voi jäädä myös rektiotäydennys, kuten esimerkissä *Hän kiisti MTV3:n uutisissa, että ero liittyisi taannoisiin kauppoihin*_{täydennys}. Esimerkit 8ä–8ö kuvastavat koodausratkaisuja, joissa *ero*-lauseke esiintyy subjektina:

8ä. *Ero* sukupuolten_{Y_ARG.} välillä **näkyi** palkkatasossa_{EI KOODATTU}. (K1560)

8ö. *Ero* ensisijaisesta hoivaajasta_{Z_ARG.} **voi aiheuttaa** lapsessa ahdistusta_{EI KOODATTU}. (K1409)

Esimerkki 8ä ilmentää intransitiivilauseetta, 8ö transitiivilauseetta. Y- ja Z-argumentit koodautuvat ”normaalisti” eli ne koodaavat *ero*-nominin täydennyksiä.

Lopuksi on syytä pohtia, mittaavatko X-, Y-, ja Z-argumentit samaa kielellistä ilmiötä kussakin *ero*-lekseemissä. Ilmeisesti mittaavat, mutta vain likimääräisesti. Semanttinen vastaavuus lienee suurinta skeeman ’ominaisuuden ja laadun samankaltaisuuden’ toteutumisissa. Argumenteissa lekseemien välinen analogia lienee suurinta elatiivisijaisella Z-argumentilla, joka spesifioi lekseemistä riippumatta tyypillisesti LÄHDEVÄYLÄ:n (Lakoff 1987: 272) eli LÄHDÄ-VÄYLÄ-KOHDE -skeeman, joka Leinon (1993: 180) mukaan on ”juurtunut erotamattomasti kognitioon ja heijastuu metaforisesti spatiaaliselta alueelta monille muille kognitiivisille alueille.” Sen sijaan ’irtisanomisen skeeman’ toteutumisissa Y-argumentit spesifioivat ainakin aloitekyvyltään erilaisia olentoja. *erota*-verbissä muutos syntyy olion aloitteesta, *erottaa*-verbissä ulkoapäin tulevasta kausatiivisesta manipulaatiosta. *ero*-nominissa aloitekyky jää usein epäselväksi. Talmyn (1985: 79–82) erottelun mukaan skeemassa EROTA.IRTI (sekä lisäksi skeemoissa EROTA.KIRKKO ja EROTA.PUOLISO) agentti (Y-argumentti) edustaa itseagentoitua kausatiivisuutta (*self-agented causative*). Tällaisessa agentissa yhdistyvät AIHEUTTAJA:n ja TEEMA:n (tai LIIKKUJA:n) semanttiset roolit – siis se, joka aiheuttaa muutoksen ja se, jonka sijaintia tai olotilan muutosta verbi predikoi (ks. esim. Langacker 1991: 283–289; Onikki-Rantajääskö 2001: 17). Skeeman EROTTAA.IRTI agentti (X-argumentti) puolestaan koodaa Talmyn kausatiivisen agentin (*causative agent*), joka kohdistaa kausaation, energiavirran, itsensä ulkopuolelle patienttiin (Y-argumentti). *erota*-verbin argumenttirakenteessa SVA_{ero}⁵⁶ (subjekti-verbi-rektioadverbiaali) S koodaa siis

⁵⁶ Notaatio A_{ero} tarkoittaa erosijaista adverbiaalitäydennystä, työssäni elatiivia.

agentin (AIHEUTTAJA ~ TEEMA ~ LIIKKUJA, Y-argumentti), V prosessin ja A rektioadverbiaalin (LÄHDE, Z-argumentti), kun taas *erottaa*-verbin argumenttirakenteessa SVOA_{ero} (subjekti-verbi-objekti-rektioadverbiaali) S koodaa agentin (AIHEUTTAJA, X-argumentti), V prosessin, O patientin (TEEMA ~ LIIKKUJA, Y-argumentti) ja A_{ero} rektioadverbiaalin (LÄHDE, Z-argumentti). *ero*-nominissa esiintyy sekä *erota*- että *erottaa*-verbin mallin mukaisista roolien jäsentymistä, ilmiö, jota sivuan analyysiluvuissa 5 ja 6.

Näiden yleisten koodauslinjausten jälkeen siirryn tarkastelemaan muuttujia, jotka mittaavat X-, Y- ja Z-argumenttien semanttista ontologiaa.

4.1.3 Muuttujien X_SEM, Y_SEM ja Z_SEM alakategoriat

X-, Y- ja Z-argumenttien semanttisia ominaisuuksia mittaavat muuttujat ja niiden luokkajako on esitetty taulukossa 10. Päädyin luokkajakoon, joka sisältää kuusi ontologista kategoriaa. Ne ovat ELOTON, KONKRETIA, INSTITUUTIO, JOUKKO, YKSIÖ ja NA.

Taulukko 10. *ero*-lekseemien täydennyksiä koskevien semanttisten muuttujien kategoriat.

Muuttujat	ID-arvot	Esimerkki aineistosta	\sum_{pesye}
X_SEM	ELOTON	vapaus, laki, muutto, VR:n toiminta, kirjavuus	1040
Y_SEM	KONKRETIA	jyvät, vaellussiika, pakkaus, vesi, thc, Gaza	300
Z_SEM	INSTITUUTIO	Kirkko, Suomi, Säätiö, kokoomus, työstä, virasta	487
	JOUKKO	köyhät, me, ne, moni	413
	YKSIÖ	sotilas, suomalainen, torvelo, Halla-Aho, hän	536

Lähtökohtanani semanttisten kategorioiden koodauksessa ovat WordNet:n (Miller 1990) 25 semanttista alkeisluokkaa (*semantic primes*), jotka toimivat lähtökohtana myös Arppella 2008. Arppen (mts. 331–334; ks. myös Arppe ja Järvikivi 2007) hengessä olen suhteellisen vapaalla kädellä muokannut alkuperäistä, englannin kielen nomineille laadittua 25 semanttisen luokan systeemiä omiin tarpeisiini sopivaksi⁵⁷. Luokitukseni ei olekaan suoraan vertailukelpoinen Arppen (2008) tai WordNet'in luokituksen kanssa. Luokittelujen taustalla ovat vaikuttaneet lisäksi – ainakin implisiittisen – Leinon (1993) ja Deshorssin (2010) työt.

Kuusi semanttista kategoriaa, joista yksi – NA – ilmoittaa että argumentti ei esiinny

⁵⁷ Muita semanttisia luokitusyhteistyöjä ovat UCREL:n (University Centre for Computer Corpus Research on Language) Semantic Annotation System (USAS), joka on tuotu myös suomen kieleen (Löfberg yms. 2003). Tunnettu nominien ontologian jaottelu on myös Lyonsin (1977: 438–460) jako ensimmäisen, toisen ja kolmannen asteen entiteetteihin, jota Karlsson yms. (1980: 71) soveltavat (ehdotuksen muodossa) suomen kieleen seuraavasti: 1) inhimillinen tai personifioitu, ts. aloitteelliseen toimintaan kykenevä olio tai sellaisten joukko, *minä, Helsingin yliopiston opiskelijat*; 2) esine tai esinejoukko, paikka: *moottoripyörä, Helsinki, työpaikkani*; 3) yhteisö, instituutio: *perhe, Keskustapuolue*; 4) aine (konkreettinen, jaollinen tarkoite): *kahvi, jauhot*; 5) teko, tapahtuma, prosessi (Lyonsin ”toinen aste”): *konsertti, esitys, näyttely*, (verbijohdokset ja *-minen*-johdokset); 6) abstrakti (Lyonsin ”kolmas aste”): *taide, tiede, rakkaus, asenne, tavoite*; i eli irrelevantti: milloin konstituentin tunnettuus merkitään irrelevantiksi.

rakenteessa, on kompromissi, johon päädyin havainnoituani aineistoani ensin introspektiivisesti. Tämän jälkeen sovelsin aineistooni aluksi hyvin hienojakoista 12 luokan jaottelua, joka jäljitteli Arppen luokitusta, sitten todella karkeaa luokitussysteemiä, nimittäin ELOLLINEN : ELOTON -jaottelua⁵⁸. Käyn luokat läpi järjestyksessä YKSILÖ, JOUKKO, INSTITUUTIO, KONKRETTIA, ELOLLINEN ja hännänhuippuna NA. Sivuan myös hienojakoisempia luokkia, jotka olen lopullisessa luokkajaossa fuusioinut isompiin luokkiin.

Luokista yksiselitteisin lienee YKSILÖ, jonka lähteenä on Arppen (2008) luokka INDIVIDUAL. Kielenaines on koodattu YKSILÖksi, kun se viittaussuhteensa nojalla poimii oliotodellisuudesta yksittäisen inhimillisen entiteetin, sellaisen kuin *johtaja*, *Halla-Aho* tai *lapsi*⁵⁹. Luokka kattaa ilmipannut persoonapronominit (*minä*, *sinä*, *hän*) ja verbeihin inkorporoidut yksikön ensimmäisen ja toisen persoonan muodot (*erosin*, *erosit*, *erotan*, *erotat*, *entajua eroa*, *et tajua eroa*) (vrt. Arppe mts. 326, alav. 115).

ero-pesyeen semantiikan erityispiirteiden vuoksi sisällytin kaksi joukkotarkoitteita mittaavaa luokkaa⁶⁰, jotka ovat JOUKKO ja INSTITUUTIO (vrt. Arppe [2008: 332] joka yhdistää nämä luokat). Rajanveto on seuraava: entiteetti saa arvokseen JOUKKO, kun kyse on useamman kuin yhden elollisen muodostamasta oliojoukosta, joka ei ole ”kovinkaan” järjestäytynyt tai ajallisesti ”pysyvä”. Tällaisia ovat *henkilöt*, *vauvat*, *sairaat*, *kaikki* ja *me*. INSTITUUTIO taas kattaa ”organisoituneemmat” joukot, joilla on ”kiteytyneempi” olemassaolon status ajassa. Näitä ovat poliittiset liikkeet ja elimet (*äärioikeisto*, *Siuntion kunnanvaltuusto*), yritykset (*Nokia*), valtiot (*Suomi*⁶¹), uskonnolliset yhteisöt (*ev.lut. kirkko*⁶²) ja esimerkiksi rikollisryhmät (*ryhmä*). Nimitys INSTITUUTIO onkin harhaanjohtava; tarkempaa olisi vaikkapa

⁵⁸ Hienojakoisen luokittelun ongelma on, että osa luokista saa vähän esiintymää, mikä on ongelma sekä khiin neliö -testeissä että klusterianalyysissä, joka on herkkä poikkeaville havainnoille. Liian karkean luokittelun ongelma on, ettei se havaintojeni mukaan selitä tutkimaani ilmiötä. Herkkyys poikkeaville havainnoille ei tosin ole yksin ongelma. Sehän tarkoittaa, että klusterianalyysi ryhmittelee ikään kuin tarkempien ohjeiden mukaan.

⁵⁹ Kokeilin testimuuttujana luokkaa PUOLISO, johon koodasin lekseemit, jotka poimivat tarkoitodellisuudesta naispuolisen entiteetin (*äiti*, *missit*) tai entiteetin, jolla viitataan avio- tai parisuhteen jäseneen (*aviomies*, *puoliso*). Koodausratkaisulla tarkastelin tämän kaltaisen sanaston kytköstä skeemoihin EROTA.PUOLISO ja ERO.PUOLISO. Vaikka luokka erotteli merkityksiä erinomaisesti poistin sen yksinkertaistaakseni luokitusta.

⁶⁰ Tämä luokkien rajanveto oli vaikea. Yksi ratkaisu olisi koodata morfologisen monikontunnisteen sisältävät yksilötarkoitteiset entiteetit YKSILÖksi. Tällöin vaikkapa sana *köyhät* koodautuisi morfologisesti monikoksi ja semanttisesti YKSILÖksi (alkeisarvo = ’köyhä ihminen’). En valinnut tätä koodaustapaa, enkä sisällyttänyt työhöni edes morfologista lukua (tai sijamuotoa) mittaavaa muuttujaa. Kokeilin tällaista muuttujaa, mutta poistin sen, koska se ei ollut merkityksiä erotteleva. Siksi työni ei tee eroa muodoille *lapio*_{KONKRETTIA} : *lapiot*_{KONKRETTIA} (huomaa kuitenkin oppostio *köyhä*_{YKSILÖ} : *köyhät*_{JOUKKO}).

⁶¹ Niin kuin Karlsson yms. (mts. 71) muistuttaa esimerkiksi sana *Valintatalo* voi tilanteesta riippuen yhtä lailla viitata liikkeeseen (LOKAATIO) kuin instituutioon (INSTITUUTIO) (vrt. Arppe 2008: 94).

⁶² En koodannut uskonnollisille kollektiiveilla (kuten *kirkko*, jolla on huomattavat 124 esiintymää aineistossa) omaa luokkaa (KIRKKO), vaikka nämä tarkoitteet kytkettyvätkin väkevästi skeemaan EROTA.KIRKKO.

organisaatioista. Luokkaan INSTITUUTIO olen analyseissä yhdistänyt lisäksi hienojakoisemman luokan SUHDE⁶³ esiintymät. Sellaisia ovat esimerkin *Fimean virkamiehillä on oikeus erota työstään* (K1052) sana *työstään* ja esimerkin *Jos ei halua muuttaa, niin erotkoon palveluksesta* sana *palveluksesta* (K1023).

Entä mihin kategoriaan eläimet kuuluvat? Olen määritellyt luokat YKSILÖ ja JOUKKO koskemaan kaikkia elollisia entiteettejä, myös eläimiä (toisin toimii Arppe [2008], jolla on erikseen luokka FAUNA). Jos eläin kuitenkin kontekstoidaan raaka-aineeksi tai ruoaksi, olen koodannut sen luokkaan KONKRETIA.

Luokkaan KONKRETIA koodasin lekseemit, joiden käsitteistyksessä nimenomaisessa puheyhteydessä etusijalla on spatiaalinen tai ”aistillinen” hahmo. Luokan mallijäseniä ovat sellaiset tarkoitteet kuin *mopo*⁶⁴, *reaktori*, *sähkötupakka*, *vesi*⁶⁵. Luokkaan kuuluu lisäksi semmoisia sanoja kuin *tissi*, *iho* ja *ruoka*, jotka Arppe (2008) on koodannut luokkiin BODY ja FOOD. Epäselvissä tilanteissa olen tapauskohtaisesti ratkaissut kieli-intuitioni nojalla, koodaanko esiintymä luokkaan KONKRETIA vai ei.

Aistillisella hahmolla tarkoitan, että olen koodannut luokkaan entiteettejä, joita ei ole pakko käsitteistää spatiaalisesti. Tällaisia ovat esimerkin *dokumenttielokuva*, *tausta* ’maalauksen taka-ala’, *seasta* ’keiton joukosta’ tai *tuosta* ’valvontakamerakuvasta’. On syytä painottaa, että koodauksessani yleisenä periaatteena ei ole ollut määrittää sanoille sanakirjamerkityksiä vaan merkitys kontekstissa. Siksi esimerkissä *Ostaja ei osaa erottaa vanhaa kalaa*-KONKRETIA *tuoreesta*-KONKRETIA (K375) myös sana *tuoreesta* on koodattu luokkaan KONKRETIA, sillä se mitä ilmeisemminkin viittaa tuoreeseen kalaan. Sanat, joilla on anaforinen viittausuhde kontekstissa aiemmin esiteltyyn spatiaaliseen entiteettiin, kuten pronominit *se*, *joka* tai *mikä* olen niin ikään koodannut luokkaan KONKRETIA (vrt. Arppe 2008: 335), mikäli korrelaattitarkoite on luokan KONKRETIA edustuma. Samaa periaatetta noudatin muiden lukkien koodauksessa. Siksi esimerkin *Tykin ammuksesta*-KONKRETIA *kumpikin*-KONKRETIA *eroaa sillä että niillä on "moottori"* (K867) sana *kumpikin* on koodattu luokkaan KONKRETIA, sillä se viittaa kontekstissa aiemmin mainittuihin sanoihin, *kathjuska* ja *missile*.

Analyseissä olen yhdistänyt KONKRETIA-luokkaan LOKAATIO-luokan lekseemit

⁶³ Tämä itse keksimäni luokan motivaationa ovat skeemojen EROTA.IRTI ja EROTTAA.IRTI yhteydessä taajaan esiintyvät rektioadverbiaalit. Luokan SUHDE voisi käsittää myös LOKAATIO:ksi (ks. Karlsson yms. 1980: 71). Luokat SUHDE (erota *töistä*) ja INSTITUUTIO (erota *hallituksesta*) yhdistyivät siksi, että merkittävä osa kummankin esiintymistä sijoittuu skeemoihin, joiden lopputunnus on IRTI.

⁶⁴ Jos lekseemin *mopo* tarkoite puheyhteydessä onkin tosiolevaisen mopon sijaan *mopo*-sana kielisysteemissä, olen koodannut esiintymän luokkaan ABSTRAKTI.

⁶⁵ Ison osan tämän luokan entiteetistä Arppe (2008) sijoittaisi hienojakoisempaan luokkaan AINE (SUBSTANCE).

hiearkkisen klusterianalyysin nojalla. Toisinkin olisi voinut menetellä⁶⁶.

Luokkaan ELTON⁶⁷ yhdistin hienojakoisemmat luokat, joiden notaatioita ovat ABSTRAKTI, LUKU, OMINAISUUS, ASIANTILA ja TOIMINTA. Luokkaan kuuluukin monenlaisia entiteettejä. Luokan ABSTRAKTI mallijäseniä ovat sellaiset sanat kuin *asia*, *käsite* tai *sana*, jotka edustavat Lyonsin (1977: 438–460) ”kolmannen tason entiteettejä”. Luokkaan olen lisäksi koodannut sellaisia sanoja kuin *internet*, *Facebook*, *kulttuuri*, *haitta*, *mielipide*, *islam*, *sukupuoli* ja *oikea*, jotka koodasin aluksi hienojakoisempiin luokkiin (kuten KOGNITIO, KOMMUNIKAATIO – erikseen sanoista, mediasta, puheaktista – ja INTERNET, joista kaksi ensimmäistä on peräisin Arppelta 2008). Luokka LUKU sisältää lukusanat, kuten *kolme*, *viides*, *ensimmäiseksi*. OMINAISUUS vastaa Arppen (mts.) luokkaa ATTRIBUTE ja sisältää sellaiset sanat kuin *kirjavuus*, *haju*, *maku* ja *hiljaisuus*.

Luokkaan ELTON liitin lisäksi hienojakoisemmat luokat, TOIMINTA ja ASIANTILA. Luokka TOIMINTA rakentuu eri tilannetyyppejä hahmottavista lekseemeistä, joita kuvaavat notaatiot TEKO, PROSESSI ja TOIMINTA (ks. Arppe 2008: 334; ks. myös VISK § 1499; Leino 1991: 158). Käsitteistyksessä on tällöin kyse spatiaalisessa dimensiossa tapahtumasta nominalistetusta, olioksi konstruoidusta prosessista, jonka semanttiseen rakenteeseen sisältyy häivytetty temporaalinen hahmo (vrt. Leino 1993: 74, 80). Luokkaan koodasin sellaisia sanoja kuin *yleisötapahtuma*, *lainmukainen toiminta*, *kuritus* ja *varastaminen*⁶⁸. Luokkaan ASIANTILA koodasin paljaat lauseet ja *se*-tukipronomini + lause -rakenteet, kuten esimerkin *Tilanne*_{TOIMINTA} *ei eroa siitä, jos kunta käy kuuntelemassa yrittäjän omakehut*_{ASIANTILA} (K944).

Erityisesti ELTON luokan entiteettien kohdalla tuli esille se, että sanojen luokittaminen ontologisiin kategorioihin on hyvin tulkinnanvaraista ja intuition ohjaamaa työtä. On usein epäselvää, mihin luokkaan sana tulisi koodata. Aidossa kielenkäytössä sanojen merkitystä voidaan manipuloida monin keinoin, esimerkiksi käyttämällä konkreettista sanaa (esi-

⁶⁶ Luokan LOKAATIO koodausratkaisuja kuvastavat esimerkit *Pk-seudun*_{LOKAATIO} *voisi toki erottaa muusta Suomesta*_{LOKAATIO} (K129) ja *Suomi*_{LOKAATIO} ja *Afganistan*_{LOKAATIO} *eroavat toisistaan kuin yö ja päivä mitä tulee geopolitiiseen sijaintiin* (K686). Edellisessä lienee kyse aluepoliittisesta irotamisesta, jälkimmäisessä INSTITUUTIO-tulkintaa vähentää ja LOKAATIO-tulkintaa vahvistaa adverbiaalilause *mitä tulee geopolitiiseen sijaintiin*.

⁶⁷ Samankaltaisen yhdistelmäluokan tekemistä ruotii Arppe (2008: 109, alav. 59) pohtiessaan vähäfrekvenssisien semanttisten kategorioiden, sellaisten kuin ATTRIBUTE, STATE tai COMMUNICATION, yhdistämistä luokkaan NOTION. Hän kuitenkin vastustaa ratkaisua: ”However, as all the semantic categories here belong to the top-level unique beginners in the WordNet ontology, one might in the resultant supersets lose in their internal coherence what one would benefit from the decrease in the number of variables.”

⁶⁸ Luokan jäsenistä suuri osa teonnimijohdoksen sisältäviä nominilausekkeita eli johdoksia, joissa johtimena esiintyy jokin seuraavista: *-minen*, *-Us*, *-ntA*, *-nti -U*, *-O*, *-UU*, *-e^x*, *-nA* (ks. VISK § 163). Luokka TOIMINTA vastaa Lyonsin (1977: 438–460) ”toisen asteen entiteettejä”.

merkiksi *kruunu*) metaforisesti (*se oli kruunu uralleni*) tai metonymisesti (*kruunu* viitattaessa kuninkaaseen)⁶⁹. Toinen tutkija olisi tehnyt paitsi erilaisen luokitusysteemin myös eri luokitusratkaisuja⁷⁰. Siksi semanttiset luokitukset ovat onnistuneimmillaankin suuntaa-antavia. Konkreettisia olioita mittaavat luokat – YKSIÖ ja JOUKKO sekä KONKRETIA (ainakin jos en olisi sisällyttänyt siihen ”aistillisia tapauksia”) – lienevät vähiten kiistanalaisia.

4.1.4 Muuttujien X_MUOTO, Y_MUOTO ja Z_MUOTO alakategoriat

X-, Y-, ja Z-argumenttien rakenteellisia ominaisuuksia mittaavat alakategoriat olen jakanut taulukon 11 mukaisesti luokkiin NP, PERS, PRON, RAKENNE JA NOLLA ja NA.

Taulukko 11. X-, Y- ja Z-argumentteja koskevat rakenteelliset muuttujat.

Muuttujat	ID-tasot	Esimerkki	\sum_{pesye}
X_MUOTO	NP	Ministeri, aineisto, yli 1000	2150
Y_MUOTO	PERS (+INKORPOROITU)	hän, me, itse, erosin	161
Z_MUOTO	PRON	Tämä, se, ne, tuo	223
	RAKENNE	Mikä on omaa ja mikä toisen omaa	158
	NOLLA (+ELLIPSI)	Konstituentti nollajäsenenä tai ellipsoituna	563
	NA	Konstituentti ”taka-alaistettuna”	1686

Toisin kuin nominilausekkeiden semanttinen luokittelu, nominilausekkeiden rakenteellisia ominaisuuksia koskevat muuttujien alakategoriat eivät ole kaikki toisensa poissulkevia. Tämä tarkoittaa, että PERS- ja PRON-luokkien jäsenet sisältyvät NP-luokkaan (yksimuuttujaisissa analyyseissä olenkin yhdistellyt luokkia). Täysin poissulkeva jaottelu voisi olla Karlssonin yms. (1980: 35) ehdottama jaottelu: ”pelkkä sana”, ”lauseke”, ”lause” ja ”0”. Toisin kuin Karlsson yms. (1980) ja Arppe (2008), joilla on käytössään isompi aineisto, olen luokittelussani pyrkinyt yksinkertaiseen jaotteluun⁷¹. En ota kantaa sanaluokkiin tai lausekkeiden jäsenten lukumäärään, vaan samaistan ”sanan” ja ”lausekkeen”⁷². Koodaustapani ei siksi tee eroa sanan *nappi*_{NP} ja keksityn lausekkeen *minun ihana nappini, jonka sain isoäidiltä*

⁶⁹ Metaforisen luennan erottaa Deshors (2010: 107), jonka muuttuja USE saa arvot METAPHORICAL ja LITERAL.

⁷⁰ Epäselviä koodattavia ovat vaikkapa sanat *kivessyöpä* ja *flunssa* esimerkissä *Sinua vastaan saattaa tulla terveyskeskuksessa ”tumpelo, joka ei erota flunssaa kivessyövästä”* (K377). Kuuluvatko ne luokkaan ABSTRAKTI? *Kivessyövällä* ja *flunssalla* on paitsi temporaalinen hahmo (koska sairaus alkaa ja ehkä loppuu) myös spatiaalinen hahmo (koska sairaus on somaattinen ilmiö). Kuuluvatko ne luokkaan TILA? Tämä luokka saisi niukasti esiintymiä. Ne on koodattu luokkaan TOIMINTA, joka analyyseissä on yhdistyy luokkaan ELOTON. Karkeampi luokitus on usein relevantimpi, koska näin vältetään hienojakoisen luokituksen tulkintaongelmat.

⁷¹ Karlssonilla yms. (1980: 34–44) muuttuja SUBJEKTIN RAKENNE sisältää 13 luokkaa: 1) rinnasteinen, 2) lause, 3) pääsana + lause, 4) infinitiivi, 5) nominaalistuma, 6) N + selittävä appositio, *kuten* tjs, 7) N + adv:li tai inf. 8) partisiippi + N, 9) attribuutti (yksi tai useampi) + N, 10) N, 11) inkorporoitu verbiin, 12) geneerinen 0, 13) ellipsi. Muuttuja OBJEKTIN RAKENNE sisältää 11 luokkaa, jotka ovat lähes samat kuin subjektilla.

⁷² Tukimuuttujat, jotka mittaavat lausekkeen jäsenten lukumäärää ja edussanan sanaluokkaa, eivät ole mukana analyyseissä. Mukana ei myöskään ole tukimuuttujat, jotka mittaavat sanajärjestystä ja argumenttien keskinäistä järjestystä suhteessa finiittiverbiin (teema- : reemapaikka -oppositio).

syntymäpäivälahjaksi sinä kesänä, kun ukista tuli lempeä_{NP} välille. Luokkajäsenyyden määrittäen lausekkeen edussanan pohjalta.

Työssäni luokkaan NP kuuluvat kaikki nominilausekkeet, joiden pääsana ei ole pronomini (tietyn ehdoin). En luokittelussa ryhmitä nominilausekkeitä alaryhmiksi, esimerkiksi infinitiivilausekkeiksi ja substantiivilausekkeiksi (ks. VISK § 491, 458, 1144). Vaikka jotkin suomen kielen pronominit heijastavat korrelaatin elollisuus : elottomuus -oppositiota, en koodannut ihmistarkoitteisia kvanttoripronomineja (sellaisia kuin *kukaan, moni, kaikki, joku*) omaan luokkaan tai luokkaan PERS – johon koodasin persoonapronominit sekä refleksiivipronominin *itse*. *itse*-sanalla kanssa tosin olisi voinut toimia muullakin tavoin.

Luokkaan INKORPOROITU sisällytin tapaukset, joissa verbin persoonapäätte on liittynyt finiittiverbiin tai kieltosanaan. Analyseissä olen tosin yleensä yhdistänyt luokat PERS ja INKORPOROITU. Ratkaisua puoltaa se, että inkorporoituun verbilausekkeeseen on aina mahdollista ilmipanna persoonapronomini, kuten lauseessa [*me*] *erosimme joella*.

Luokkaan PRON koodasin anaforiset demonstratiivipronominit (kuten *tämä, tuo* ja *se*), joilla viitataan tilanteisiin ja käsitteisiin mutta myös ihmisiin (ks. VISK § 720–740). Luokkaan koodasin myös relatiivipronominit (*joka* ja *mikä*), joille loin aluksi oman luokan, jonka poistin esiintymien vähyyden vuoksi. Esimerkit 9a–9c ilmentävät koodausratkaisuja:

9a. Mitä tarkoittaa yksikäsitteisesti hyväksyminen? Sitä, että nakkivarkaat teloitetaan? Miten tämä_{PRON} eroaa jossain tilanteessa hyväksymisestä?_{NP} (K943)

9b. Ei se_{PRON} [saksalainen junatyypin] eroa mitenkään muista junista_{NP}. (K639)

9c. Se pieni alkupääoma ja tietämys joka_{PRON} erottaa köyhät_{NP} ja keskiluokkaiset_{NP}. (K445)

PRON-luokan ytimen muodostavat juurikin *tämä ~ tuo*- (9a), *se ~ ne*- (9b) ja *joka*-lausekkeet (9c). NP:ksi olen kuitenkin koodannut lausekkeet, joissa *se*-lekseemi esiintyy demonstratiivisena määritteenä tai tarkenteena (ks. VISK § 564), kuten esimerkissä *Siitä pahasta, joka tappaa_{NP}* ja *Jos siitä kirkosta_{NP} eroaisi*.

”Lausetta” muuttujien alakategorioista edustaa RAKENNE, jota havainnollistaa esimerkit 9d–9f:

9d. Miten tu_{OPRON} eroaisi siitä, että väittäisin merenpinnan nousevan_{RAKENNE} (K874)

9e. Milläs erotat_{INKORPOROITU}, mikä rangaistus on nöyryyttävä_{RAKENNE} ja mikä ei_{RAKENNE} (K305)

9f. Jos henkilö_{NP} ei erota, että olisi mitään eroa paloitella veitsellä koira_{RAKENNE} tai porkkan_{RAKENNE}, niin [..] (K307)

Esimerkissä 9d Z-argumentin spesifioi *se*-tukipronomini + lause -rakenne. Y-argumentin luokkana on PRON. 9e:ssä X-argumentin luokka INKORPOROITU määräytyy verbin

persoonaliitteen nojalla. Y- ja Z-argumentin spesifioivat epäsuoran kysymyslauseen rinnasteiset osat. 9f:ssä Y- ja Z-argumentin spesifioivat paljaan *että*-lauseen rinnasteiset osat, *että olisi eroa paloitella veitsellä koira* ja [*että olisi eroa paloitella veitsellä*] *porkkana*.

Vielä tarkastelematta ovat tapaukset, joissa X-, Y- tai Z-argumentti puuttuu lauserakenteesta eli luokat NOLLA ja NA. Lisäksi muuttujistossani oli luokka ELLIPSI, jonka pienen esiintymistajuuden vuoksi yhdistin luokkaan NOLLA, olkoonkin, että kyse on eri kielenilmiöistä⁷³. En usko että luokkien yhdistäminen vääristää tuloksia.

NOLLA:n olen määritellyt siten, että kyse on semanttisen valenssin nojalla ilmauksen merkitysrakenteeseen kuuluvasta jäsenestä, joka on läsnä käsitteistyksessä mutta jolta puuttuu ilmirakenteen koodaus. Nolla voi esiintyä kaikkien argumenttien paikoilla, mutta yleisin se on subjektin eli X- ja Y-argumenttien paikoilla. Kun nollasubjekti on ihmisviitteinen on kyse *nollapersoonasta* eli ilmiänsä olemattomasta ”avoimesta” NP:stä, joka voi viitata keneen tahansa ihmiseen ja usein nimenomaan puhujaan itseensä tai puhuteltavaan (ks. VISK § 1345–1347)⁷⁴. *erota*-verbin NOLLA:ksi koodattuja Y-argumentteja edustavat esimerkit 9g–i:

9g. Tiedän naisia, jotka sietävät väkivaltaa suhteessa. Yksinäisyys taloudellinen turvattomuus, muutoksen pelko, (kerran jopa ahdistus siitä, kuka vaihtaa talvirenkaat autoon jos [NOLLA] **eroaa**!) (K977)

9h. Ja höpö, höpö Wallin! Ei [NOLLA] kannata selitellä vaan **erota**! (K1068)

9i. Jos häpeää sitä puoluetta, johon on liittynyt, [NOLLA] **voi** siitä aina **erota**. (K518)

Esimerkissä 9g nolla viittaa paitsi naisiin yleensä eritoten tiettyyn naiseen, jolle ahdistus auton talvirenkaista nousee seinäksi eron eteen. Esimerkki 9h edustaa ellipsiä, jossa Y-argumentin spesifioiva *Wallin* on jätetty toistamatta. Y-argumentin spesifioivista nolista esimerkin 9i tavoin modaalisen apuverbin sisältää 70.3 %. Ne on koodattu yleensä nolliksi.

Luokan NA olen määritellyt siten, että kyse on elementistä, jota semanttinen valenssi ei edellytä käsitteistettäväksi tai elementistä, jonka käsitteistäminen on ”taka-alaistettu”. *erota*-verbin yhteydessä Y-argumentti koodautuu NA:ksi seuraavan tapaisissa esimerkeissä:

9j. Länsimaissa **erotaan** eikä enää uskota, Afrikassa käännytään. (K744)

9k. Lapsena koettu vanhempien avioero [...] lisää todennäköisyyttä **erota**. (K987)

⁷³ Ellipsillä tarkoitetaan että kahdella perättäisellä lauseella tai lauseenosalla on yhteinen jäsen (ks. VISK § 1177). Nollapersoonassa taas on kyse epämääräisestä persoonasta (Setälä 1880: 96). Lisäksi puhutaan nolla-anaforasta, jossa edellä puheena olleeseen tarkoitteeseen viitataan käyttämättä anaforista pronominia (VISK § 1431). Aineistoni muutamat ilmeiset nolla-anaforat olen koodannut luokkaan NP, kuten *Sitä Sergei ei tehnyt, vaan [NP] **erotti** Halosen* (K345).

⁷⁴ En ota kantaa siihen, onko nollan tulkinta geneerinen vai puhujaan tai puhuteltavaa koskeva. Rajasin työstä ulos korpusaineistolle koodaamani, tunnettuutta mittaavat muuttujat X_SPESIES, Y_SPESIES ja Z_SPESIES, joiden ala-arvot ovat GENEERINEN, INDEFINIITTINEN ja DEFINIITTINEN (vrt. Vilksen 1980).

Nämä ovat subjektittomia rakenteita: esimerkissä 9j *erota*-verbi on yksipersonaisessa passiivissa, kun taas 9k:ssä *erota*-verbi toimii *todennäköisyys*-sanana infiniittimäärityksenä. Kaikissa *erota*-verbin havainnoissa X-argumentti koodautuu luokkaan NA.

erottaa-verbin ja transitiivilauseessa objektina esiintyvän *ero*-nominin nollapersoonat ovat X-argumentteja. Niitä hahmottavat seuraavat esimerkit:

9l. Jollei [NOLLA] **osaa** ampua saati **erota**, mitä on ampumassa, [...] jättäköön homman paremmalleen. (K237)

9m. Jos tuota **eroa** eri kanavien välillä [NOLLA] **ei tajua**, niin mainoksia tunkee [...] niin paljon, että ehtii käydä jääkaapilla tai tupakilla välillä ja mainokset vaan jatkuvat. (K1089)

Esimerkissä 9l nollan tilalle voidaan käsitteistää joko kontekstissa aikaisemmin mainittu metsästäjä eli 'definiittinen metsästäjä' tai geneerisesti 'kuka tahansa metsästäjä'. Esimerkissä 9m ilmeisin nollan paikalle sijoittuva olio on kontekstin perusteella puhuteltava.

Epäselvemmin koodattavia ovat nesessiivisyyttä ilmaisevan modaalisen apuverbin sisältävät subjektittomat lauseet. Karlsson yms. (1980: 28) kirjoittaa: "Mikäli yksipersonaiset modaaliset apuverbit *täytyy*, *pitää*, *kannattaa*, jotka voivat saada vain genetiivisubjektin, ovat ilman subjektia, on vaikea sanoa, onko kysymys geneerisestä nolla-subjektista vai passiivista. Tällaiset tapaukset on aina kooditettu passiiviksi." VISK (§ 1355) taas sanoo *pitää*-verbin käytöstä seuraavasti: "Koska nesessiivirakenteesta ei voi muodostaa yksipersonaista passiivia, nollasubjektillinen nesessiivirakenne on hahmotettavissa tätä aukkoa korvaavaksi. Nollapersoonan ja passiivin ero siis kumoutuu nesessiivirakenteissa." Seuraavat esimerkkien X-argumentit onkin koodattu luokkaan NA, sillä kyse on argumentista, jonka käsitteistäminen on "taka-alastettu".

9n. Nurin menneet varmaankin [NA] **erotettiin** jengistä. (K21)

9o. Kirkko ja yliopisto [NA] **pitää** heti **erottaa** toisistaan. (K8)

9p. Sittenhän voitaisiin järjestää kansanäänestys siitä, [NA] **tuleeko** hallitus **erottaa**. (K84)

Esimerkki 9n edustaa yksipersonaista passiivia, joka ilmentää "prototyypisintä" X-argumentin NA-koodausesimerkkiä – taka-alastettua, indefiniittistä agenttia. Esimerkki 9o havainnollistaa aineistossani yleistä nesessiivistä⁷⁵ konstruktia, johon sopii Karlssonin sanat: "ei juurikaan poikkeaa passiivista". Näyttämömetaforaa käyttäen agentti on tällöin häivytetty verhojen taakse – ehkä jopa selvemmin kuin muodollisissa passiivilauseissa! Aineistossani tavallista on, että yksipersonaista verbiä edeltävien tai niiden perässä tulevien lauseiden verbit ovat passiivissa, kuten esimerkin 9p *voitaisiin*. Modaaliverbi voi tosin olla

⁷⁵ *erota*-verbissä nesessiivisyys tuottaa yleensä NOLLA-luennan, kuten esimerkissä *Ellei viihdy liitossaan, joko pitää erota tai sitten kärsiä hiljaa* (K988).

mahdollisuuttakin ilmaiseva niin kuin esimerkissä *Kiusaajaa ei* [NA] *voi erottaa*, koska silloin *mietitään* mitä hänelle käy (K42). Minimipari *Kiusaajaa ei voi erottaa* : *Kiusaajaa ei voida erottaa*, ilmentää VISK:n (§ 1363) sanoja siitä, että *voida*-verbiä käytettäessä ”on monesti hyvin vähän eroa passiivin ja nollapersoonan välillä”. Luokkaan NA olen erikoissyistä joskus koodannut *on vaikeaa* -rakenteen toteutumat:

9q. Sasi myönsi, että [NA] **olisi järkevää erottaa** Kreikka ja että sinne syydetään rahaa hukkaan. (K85)

9r. Se on saanut virheellisen nimen maamyyrä siksi, että vain harvat ovat sen nähneet ja sen multakekoja **on** [NOLLA:n] **vaikea erottaa** oja- eli vesimyyrien tekemistä. (K228)

Esimerkin 9q X-argumentti on koodattu luokkaan NA, koska kontekstissa esiintyy passiivia (*syydetään*) ja koska taho, joka erottaisi Kreikan, on lauseyhteyden perusteella indefiniittinen (peruspuhuja tuskin edes tietää, miten EU-maa irtisanotaan Euroopan unionista). Esimerkki 9r on koodattu NOLLA:ksi, koska se mahdollistaa sekä definiittisen että geneerisen tulkinnan⁷⁶. Kontekstissa ei nyt esiinny passiivia. Sen sijaan rinnasteisen *että*-lauseen alkuosassa on ilmipantu NP *harvat*, joka niin sanoakseni kutsuu lukijaa astumaan nollan paikalle (vrt. Laitinen 1995).

ero-nominin sisältävissä intransitiivi- ja kopulalauseissa Y-argumenttien ja Z-argumenttien nollaedustusta hahmottavat esimerkit 9s–9u:

9s. Kun/jos valtio ei käytä leikkuria täysillä, kuntien on se tehtävä. Tulos kansalaisen kannalta täsmälleen sama. [NOLLA:n ja NOLLA:n] Ainoa **ero on**, että [...] (K1335)

9t. [NOLLA:n] **Ero** korkeisiin töyssyihin_{Z_ARG}. **on** se, että [...] (K1337)

9u. Ainoa **ero** vertailuryhmillä_{Y_ARG}. **oli** ero tietokone- ja konsopelien pelaamisessa? (K1622)

Esimerkissä 9s sekä Y- että Z-argumentit on koodattu NOLLA:ksi, 9t:ssä vain Y-argumentti. Esimerkissä 9u NOLLA:ksi koodautuu Z-argumentti. Samalla tavalla olen koodannut muuttujan RAKENNE luokan Y_JAOLLINEN esiintymät myös *erottaa*- ja *erota*-verbin sisältävissä esimerkeissä, kuten esimerkissä *Kaikki parit eroavat ja pojilla ei ole miehen mallia* (K984). NA-tapauksia havainnollistavat *ero*-nominin sisältävät esimerkit *Silloin ero on paras ratkaisu* (K1642) ja *Vanhempieni ero on parasta* (K1614), joista edeltävässä sekä Y- että Z-argumentit on koodattu NA:ksi ja jälkimmäisessä vain Z-argumentti.

Z-argumentti koodautuu NA:ksi *erota*- ja *erottaa*-verbeissä, kun LÄHTEEN käsitteistäminen taka-alaistuu ja NOLLA:ksi, kun kyse on ellipsiin rinnastettavasta tilanteesta. Tällaisia koodaustilanteita hahmottavat nämä esimerkit:

⁷⁶ Agentin luokkajako NA : NOLLA : NP heijastelee karkeasti jatkumoa, joka alkaa indefiniittisestä agentista (NA passiivin vastineena), jatkuu kontekstin pohjalta definiittiseen agenttiin (NOLLA puhujaan tai puhuteltavaan viittaavana) tai geneeriseen agenttiin (NOLLA geneerisenä) ja päättyy definiittiseen, ilmipantuun agenttiin (NP).

9v. Asuntoministeri Kiurun_{Y_ARG}. **tulisi erota**. (K1067)

9x. Viisas optiojohto **erotti** kaikki linjamiehet_{Y_ARG..} (K433)

9y. Piispat ovat [...] vaatineet Luther-säätiötä eroamaan kirkosta. Mutta miksi sellainen säätiö_{Y_ARG..} **eroaisi** [virallisesta kirkosta], joka tulkitsee olevansa oikea oppisempi. (K897)

9z. Itse erotan hajun perusteella tupakan sähkötupakasta satavarmasti. Jos hajuaistissa ei ole vikaa, niin [tupakan sähkötupakasta] **erottaa** myös taksikuski.

Esimerkeissä 9v ja 9x on kyse lauseista, joista Z-argumentin spesifioimat oliot ('ministerin työ' ja 'hallitus') ovat mukana käsitteistyksen merkitystaustassa, mutta leksikaalisen ilmipanemattomuuden takia taka-alastetusti (NA). Esimerkit 9y ja 9z ilmoittavat ellipsille ominaista tilannetta: 9y:ssä NOLLA:ksi koodautuu Y-argumentti, 9z:ssä myös Z-argumentti.

Lopuksi sopii pohtia, mittaavatko NA ja NOLLA *ero*-lekseemien välillä samaa kielellistä ilmiötä. Ilmeisesti eivät, ja ongelmia on monia. Y-argumentteina esiintyvät NOLLA:t ilmoittavat *erota*-verbissä yleensä nolla-agenttia. *erottaa*-verbin Y-argumentti on ilmipanematta harvoin, jolloin se signaloi ellipsiä, joka on koodattu NOLLA:ksi. Muuttujan RAKENNE luokkaan Y_JAOILLINEN kuuluvissa lauseissa Z-argumentti on koodattu NA:ksi, koska sen tarkoite on spesifioitu jaollistarkoitteisessa Y-argumentissa. Z-argumentti on koodattu NA:ksi myös lauseissa, joissa lähde on oikeasti jätetty ilmipanematta rakenteeseen. X-argumentin olen koodannut NA:ksi *erottaa*-verbin sisältävissä subjektittomissa lauseissa kuten passiivissa, mutta myös *ero*-nominillisissa kopulalauseissa ja *erota*-verbin sisältävissä intransitiivilauseissa, vaikkei näiden merkitysrakenteeseen kuulu lainkaan tekoketjun kausatiivista jäsentä, joka olisi voitu "taka-alastaa". Muun muassa näiden ongelmien vuoksi olen analyysissä purkanut NOLLA : NA -oppositiota (etenkin muuttujassa Y_MUOTO) fuusioimalla nämä luokat. Lisäksi monimuuttujaisessa analyysissä käytän tukimuuttujaa LUOKKA ilmoittamaan, onko kyse intransitiivilauseesta, transitiivilauseesta vai kopulalauseesta.

4.2 Argumenttirakennetta koskevat muuttujat

Y- ja Z-argumenttien jäsentymistä *ero*-lausekkeissa mittaa kaksi muuttujaa, jotka ovat RAKENNE ja RESIPROOKKI. Tarkastelen niitä luvuissa 4.2.1 ja 4.2.2. Muuttujat mittaavat osittain samaa ilmiötä kuin Y-muoto ja Z-muoto – asia, johon palaan analyysissä (ks. luku 5) ja jonka vuoksi muuttuja RAKENNE on poistettu monimuuttujaisesta analyysistä (ks. luku 6).

4.2.1 RAKENNE

Muuttujan RAKENNE alaluokat ovat VAIN_Y, Y_JAOLLINEN, KONNEKTORI, REKTIO, REKTIO2, VAIN_Z ja NA. Muuttuja ei ota kantaa siihen, esiintyykö rakenteessa ehkä lisäksi X-argumentti, RESIPROOKKI tai teemapaikalla kehysadverbiaali, vaan mittaa ainoastaan Y- ja Z-argumenttien statusta. Muuttuja ei huomio, missä sijamuodossa argumentti esiintyvät. Taulukko 12 esittää muuttujan RAKENNE alaluokitukset.

Taulukko 12. Muuttujan RAKENNE luokkajako.

Muuttuja	ID-tasot	Esimerkki	\sum_{pesye}
RAKENNE	VAIN_Y	<u>Vanhanen</u> erosi .	233
	Y_JAOLLINEN	Kun tuttu rekisteröitynyt <u>lesbopari</u> erosi .	222
	KONNEKTORI	Oppii erottamaan <u>ystävät ja viholliset</u> .	378
	REKTIO	Miten <u>moraali</u> eroaa laista?	537
	REKTIO2	<u>Veikkauksen lotossa</u> on eroa kilpailijoihinsa.	58
	VAIN_Z	<u>Ay-liikkeestä</u> voi erota , kun ilmoittaa työnantajalle.	73
	NA	Erotaan ystävinä.	164

Muuttujan luokaksi koodasin VAIN_Y silloin, kun *ero*-lausekkeessa täydennyksistä ilmipantuna on vain Y-argumentti. Muuttuja mittaa tällöin samalla Z-argumentin poissaoloa. Aineistoesimerkit 10a–10c edustavat luokan VAIN_Y esiintymiä eri *ero*-lekseemeissä.

10a. Suomen_{Y_ARG}. **ero** ensiksi **auttaisi** myös muita euromaita. (K1566)

10b. Vanhanen_{Y_ARG}. **erosi**. (K1008)

10c. Suomessa **ei** Virkamiestä_{Y_ARG}. **voi erottaa**. (K29)

Esimerkissä 10a Y-argumentiksi koodautuu *ero*-lekseemin ensimmäinen täydennys *Suomen*, 10b:ssä subjektikonstituentti *Vanhanen* ja 10c:ssä *Virkamiestä*. Luokan VAIN_Y eri-koistapausta edustaa luokka Y_JAOLLINEN, jonka koodausratkaisuja havainnollistavat esimerkit 10d ja 10e:

10d. Kun tuttu rekisteröitynyt lesbopari_{Y_ARG}. **erosi**, totesi isä irvistäen, että [...]. (K968)

10e. Ihmiisiä_{Y_ARG}. **yhdistävät tai erottavat** muut(kin) asiat kuin kansallisuus. (K460)

Luokkaan Y_JAOLLINEN olen koodannut esimerkit, joissa Y-argumentin spesifioi kollektiivisana (ks. VISK § 557), kuten esimerkin 10d *lesbopari*, sekä esimerkit, joissa Y-argumentin spesifioima elementti on morfologisesti monikossa (10e), toisin sanoen, kun Y-argumenttina esiintyy lauseke, joka kattaa semantisesti enemmän kuin kaksi entiteettiä. En erota kokonaisuutena (*lesbopari*) tai distributiivisesti (*kansa*) hahmotettavia kollektiivityyppejä, vaan samaistan kaikki sanat, jotka on ontologisesti purettavissa useaan osatekijöihin. Ainesanat, kuten *uraani*, eivät kuitenkaan laukaise Y_JAOLLINEN-tulkintaa. Tyypillistä on, että Y_JAOLLINEN-luokan yhteydessä esiintyvän Y-argumentin spesifikaation minimipariksi on

periaatteessa ajateltavissa konstruktio, jossa jaollinen entiteetti ilmaistaan useamman konstituentin avulla. Esimerkin 10d minimipariksi voidaan muokata lause *Kun tuttu lesbo_{Y_ARG}. erosi puolisostaan ~ toisesta lesbosta_{Z_ARG}*, joka olisi luokan REKTIO toteutuma (ks. s. 63).

Luokkaan KONNEKTORI olen koodannut aineistoesimerkit, joissa ilmipannut Y- ja Z-argumentit esiintyvät ”konnektorilla” (tyypillisesti *ja*-, *vai*- tai *tai*-rinnastuskonjunktiolla) toisiinsa kytkettyinä. *erota*- ja *erottaa*-verbillä luokkaan kuulumisen kriteerinä on, ettei Z-argumentti esiinny rektiosijoissa (elatiivi ~ illatiivi). Luokkaan KONNEKTORI sisältyy aineistohavainnoista *ero*-nominilla 42.9 %, *erota*-verbillä 3.6 % ja *erottaa*-verbillä 19.4 %. Koodausratkaisuja demonstroivat esimerkit 10f–10i:

10f. Viimeiset 15 vuotta CO2_{Y_ARG}. ja konnektori lämpötila_{Z_ARG}. **ovat eronneet** toisistaan. (K569)

10g. **Oliko** kyselyssä **erotettu** romaanikerjäläiset_{Y_ARG}. ja konnektori romaanit?_{Z_ARG}. (K137)

10h. Ensisijaisen_{Y_ARG}. ja konnektori toissijaisen perusturvan tasoissa_{Z_ARG}. **ei ole eroa**.⁷⁷ (K1110)

10i. Siinä on merkittävä **ero** onko skanneri mukana kuljetettava_{Y_ARG}. vai konnektori kiinteä_{Z_ARG}.⁷⁸ (K1552)

Esimerkissä 10f luokkajäsenyyden määrittää rinnasteinen subjekti *CO2 ja lämpötila*, 10g:ssä rinnasteinen objekti *romaanikerjäläiset ja romaanit*, esimerkissä 10h *ero*-nominin täydennykset ja 10i:ssä rinnasteinen lausetäydennys. *ero*-nominilla Y- ja Z-argumentit esiintyvät seuraavissa sijamuotokombinaatioissa: genetiivi + genetiivi (n=134)⁷⁹, olosija + olosija (n=74), olosija (n=16), lause (n=14), genetiivi + allatiivi (n=6), nominatiivi + nominatiivi (n=2), illatiivi + illatiivi (n=1), allatiivi (n=1), elatiivi (n=1). Sijat jakautuvat epätasaisesti *ero*- ja *eroa*-muotojen välille.

Luokkaan REKTIO olen koodannut rakenteet, joissa ilmipannun Y-argumentin lisäksi esiintyy erosijainen Z-argumentti, kuten esimerkeissä 10j–10k:

10j. Miten sikala_{Y_ARG}. **eroaa** turkistarhasta?_{Z_ARG}. (K939)

⁷⁷ Lausekkeet, kuten *ensisijaisen ja toissijaisen perusturvan tasoissa*, jotka edustavat kaavaa [[Y_arg.genetiivi + Z_arg.genetiivi] FOKUS_{olosija}], olen hahmottanut siten, että Y-argumentti käsittää lausekkeen *ensisijaisen perusturvan tasossa* ja Z-argumentti lausekkeen *toissijaisen perusturvan tasossa*. Vaihtoehtoinen koodausratkaisu olisi ollut koodata lauseke *perusturvan tasoissa* muuttujan FOKUS esiintymäksi. Tällöin Y-argumenttina toimisi lauseke *ensisijainen perusturva* ja Z-argumenttina *toissijainen perusturva*. Jälkimmäinen ratkaisu on perusteltavissa sillä, että muuttujistoni sisältää muuttujan nimeltään FOKUS, valitsemani taas sillä, että inessiivisijainen konstituentti toimii lausekkeen edussanana, jota genetiivisijaiset konstituentit määrittävät.

⁷⁸ Periaatteessa nämä rakenteet olisi voinut koodata myös muuttujan VASTAUS esiintymiksi. Tämän kaltaisissa esimerkeissä (n=9) *siinä ~ sillä* pronomini viittaakin samalla (ainakin implisiittisesti) paitsi aiemmin mainittuun kahden entiteetin vertailuun (Y- ja Z-argumentit), myös niiden väliseen eroavaisuuteen (muuttuja VASTAUS). Esimerkki 10i on mahdollista editoida muotoon: *Eri skannereiden välillä on merkittävä ero siinä, onko skanneri kuljetettava vai kiinteä*.

⁷⁹ Kombinaation Y_arg.genetiivi + Z_arg.genetiivi esiintymistä 75 on muotoa Y_arg.genetiivi + Z_arg.genetiivi + välillä, kuten esimerkissä *kas kun seksin ja rakastelun välillä on iso ero* (K1541).

10k. Nuoret_{X_ARG}. itse **erottavat** hyvin kiusaamisen_{Y_ARG}. naljailusta._{Z_ARG}. (K417)

10l. Kuka **on ollut vaatimassa** kaikkien_{Y_ARG}. **eroa** kirkosta?_{Z_ARG}. (K1042)

erota-verbin yhteydessä (10j) argumenttirakenteena toimii SVA_{ero}, kun taas *erottaa*-verbin yhteydessä (10k) argumenttirakenteena toimii (S)VOA_{ero}. Luokkaan sijoittuu *erota*-verbillä 51.2 %, *erottaa*-verbillä⁸⁰ 47.7 % ja *ero*-nominilla 1.1 %, havainnoista. Huomion-arvoista on, että *ero*-nominilla luokalla REKTIO on vain 7 esiintymää, kuten 10l, jossa ensimmäinen täydennys (*kaikkien*) on genetiivissä. *ero*-nomini painottuikin enemmän luokkaan REKTIO2 (n=56), joka vastaa täydennysten yhteydessä esiteltyä tyyppiä 3 (ks. luku 4.1.2). Sen tunniste on illatiivisijainen Z-argumentti, jonka spesifioi esimerkissä *Mikä on ero tähän tapahtumaan* lauseke *tähän tapahtumaan*. Luokan REKTIO2 koodauksessa poikkesin koodauslinjasta, koska koodasin luokkaan myös havainnot, joissa Y-argumentti on poissa (n=30), toisin sanoen, esiintyy kontekstissa ilmipantuna ennen *ero*-lauseketta. Pesyeen klusterianalyseissä olen yhdistänyt luokat REKTIO2 ja REKTIO, koska kummassakin Y- ja Z-argumentti rinnastetaan, ei ”konnektorin”, vaan sijamuotojen avulla.

Lisäksi loin luokan VAIN_Z. Tällöin ilmipantuna on vain elatiivisijainen Z-argumentti, kuten esimerkissä 10m–10n.

10m. **Voiko** Ay-liikkeestä_{Z_ARG}. **erota** samalla tavalla? (K619)

10n. Jos eurosta eroaminen ~~ei~~ koodattu **merkitsee eroa** myös EU:sta [...] (K1250)

Luokan VAIN_Z toteutumat (n=71) painottuvat *erota*-verbiin (n=65), jossa ne ovat joko nollasubjektillisia lauseita (n=49), yksipersoonaisia passiiveja (n=10) tai muita subjektittomia rakenteita (n=6), kuten kysymyslauseita (10m). *erottaa*-verbillä luokalla on 2 esiintymää ja *ero*-nominilla 6. Analyyseissä olen usein yhdistänyt luokat REKTIO ja VAIN_Z.

Vielä esittelemättä on luokka, jonka tunnus on NA. Koodasin NA:ksi aineistoesimerkit, joissa *ero*-lauseke esiintyy ilman Y- tai Z-täydennystä, kuten esimerkeissä 10o–10r:

10o. Silloin **ero on** paras ratkaisu. (K1642)

10p. NykYTEknologialla on kahdet kasvot. Se yhdistää ja **erottaa**. (K450)

10q. **Erotaan** ystävinä. (K20)

⁸⁰ Koodasin luokkaan REKTIO kolme *erottaa*-verbin havaintoa, joissa Z-argumentti esiintyy tulosijassa. Yksi niistä on esimerkki *Erotetaan lappilaiset ja saamelaiset suomalaisen*_{Y_ARG}. *yhteiskunnan tukien ulkopuolelle*_{Z_ARG} (K164). Luokkaan KONNEKTORI koodasin viisi *erottaa*-verbin havaintoa, joissa esiintyy tulosijainen adverbiaali. Yksi niistä on esimerkki *Kannattaa erottaa Gmailin*_{Y_ARG}. *ja konnektori Googlen hakujen käyttöä*_{Z_ARG}. *eri selaimiin*_{TULORAKENNE} (K168). Yhden tulosijaisen esimerkin koodasin luokkaan Y_JAOLLINEN. Luokkaan REKTIO koodasin myös esimerkit joissa esiintyy sekä *ero*- että tulosijainen adverbiaali, kuten esimerkissä *Alueen partiosta*_{Z_ARG}. *on hankala erottaa miestä*_{Y_ARG}. *liikenteenohjaukseen*_{TULORAKENNE} (K45). Aineistossani tulosijainen adverbiaali spesifioi Z-argumentin 3 tapauksessa 13:sta. 10 kertaa tulosijainen adverbiaali koodautuu luokkaan TULORAKENNE, joka ei ole mukana analyyseissäni.

10r. Toisessa olen kirjoilla, maksavana jäsenenä. Siksi kai Kirkossa käyn hieman useammin kuin ravintolassa. Siinäkin eroavat. (K818)

Kuten esimerkeistä ilmenee, *ero*-lekseemit eivät aivan yksin esiinny. Esimerkissä 10o läsnä on predikatiivi (*paras ratkaisu*), 10p:ssä X-argumentti eli subjekti (*se*), 10q:ssä essiivisijainen adverbiaali (*ystävänä*) ja 10r:ssä VASTAUS-muuttujan spesifioiva adverbialitädennys (*siinäkin*), joista yksikään ei kuitenkaan ole Y- tai Z-argumentti. Luokkaan NA sijoittuu *ero*-nominilla 21.4 %, *erottaa*-verbillä 1.4 % ja *erota*-verbillä 5.7 % havainnoista. Ominaisin luokka NA on *ero*-nominille, ja siinä skeemoille ERO.PUOLISO, ERO.NUMEROT ja ERO.ERI, jotka esiintyvät usein ilman täydennyksiä tai siten, että täydennykset on mainittu kontekstissa aikaisemmin. NA-luokan iso edustus *ero*-nominissa johtuukin periaatteestani koodata vain lauserajan (tai maksimissaan virkerajan) sisällä olevia kielenelementtejä.

4.2.2 RESIPROOKKI

Karsiakseni muuttujan RAKENNE luokkia loin oman muuttujan resiprookista eli vastavuoroista suhdetta ilmaiseville kielenelementeille. Muuttujan RESIPROOKKI ja RAKENNE kytköksiä tarkastelen analyysissä (luvussa 5). ”Resiprookkisuus” on *ero*-pesyeen luontainen ominaisuus. VISK:n (§ 486) mukaan ”resiprookkiset verbit [joihin VISK lukee verbit *erota* ja *erottaa*] edellyttävät toisiinsa suuntautuvia osallistujia, joiden kesken jotakin tapahtuu. Niillä on systemaattista täydennysten muodon vaihtelua. Toimijat voidaan ilmaista joko joukkotarkoitteisella, ts. monikollisella [*Kissat eroavat toisistaan*] tai rinnasteisella subjektilla [*Y ja Z eroavat*], tai sitten kahtena lausekkeena, joista toinen on subjekti ja toinen *kanssa*-lauseke [*Y ja Z eroavat toisistaan*].”

VISK jatkaa, että ”[...] eroamista ilmaisevien intransitiiviverbien yhteydessä toinen täydennys on suuntasijainen adverbiaali; [transitiivilauseessa] sama vaihtelu koskee objektia.” Havaintona vaihtelusta VISK esittää minimiparin *A ja B eroavat (toisistaan) ~ A eroaa B:stä*, jota VISK tulkitsee siten, että ”eri ilmaisuvaihtoehtojen käyttö riippuu siitä, esittääkö osallistujat tasa-arvoisina vai tarkastellaanko asiaa yhden näkökulmasta”.

Palaan aiheeseen vielä luvun lopuksi ja tarkastelen nyt muuttujan RESIPROOKKI koodausta. Taulukko 13 esittää muuttujan luokkajako:

Taulukko 13. Muuttujan RESIPROOKKI luokkajako.

Muuttuja	ID-tasot	Esimerkki	\sum_{pesye}
RESIPROOKKI	LÄSNÄ	<i>toisistaan, välillä, kanssa</i>	222
	NA	ei esiinny lauserakenteessa	1444

Kuten taulukosta 13 ilmenee, muuttujan RESIPROOKKI luokkajako on dikotominen: resiprookkinen elementti on rakenteessa ilmipantuna (LÄSNÄ) tai sitten ei ole, jolloin muuttuja

saa arvokseen NA. Esimerkit 11a–11b havainnoivat *erota-* (11a) tai *erottaa-*verbin (11b) sisältäviä esimerkkejä, joissa muuttujan RESIPROOKKI arvoksi on koodattu LÄSNÄ:

11a. Ei ampuminen kovilla ja pehmeillä kuulilla niin paljon **eroa** toisistaan (K648)

11b. Harvempi tavallisista tallajista **erottaa** ykköstä ja nelosta toisistaan. (K390)

*ero-*verbien yhteydessä muuttujan RESIPROOKKI spesifioi sana *toisistaan* (10a–10b), joka on sanasta *toinen* (kaavalla *toinen* + monikon *i* + elatiivin *sta* + 3. persoonan possessiivisuffiksi *-Vn*) muodostettu resiprookkipronomini (ks. VISK § 733, 95). *ero-*nominissa muuttujan RESIPROOKKI spesifikaatioina toimivat sellaiset sanat kuin *kanssa* (n=1), *kesken* (n=3), *versus* (n=1), *vs* (n=1), *'/'* (n=1), *välillä* (n=89), *välille* (n=7), *väliltä* (n=1), *välistä* (n=1), *välinen* (n=3), *välisen* (n=1) ja *välisessä* (n=1)⁸¹. Koodasin ne luokkaan RESIPROOKKI, sillä nähdäkseni niillä on *ero-*nominin yhteydessä samankaltainen kielellinen funktio kuin *toisistaan*-lekseemillä *ero-*verbien yhteydessä. Syntaktisesti ne poikkeavat siinä, että *välillä*-sana on *ero-*nominin täydennyslausekkeen edussanana, kun taas *toisistaan*-sana esiintyy rek-tioadverbiaalin paikalla, eikä ole kielipollisesti pakollinen. *ero-*nominin koodausratkaisuja havainnollistavat seuraavat esimerkit:

10c. Oleellisin **ero** Haaviston ja Niinistön välillä **on** seinien sensuurissa. (K1417)

10d. **On tärkeää tehdä ero** vapaaehtoisen ja pakotetun seksityön välille. (K1615)

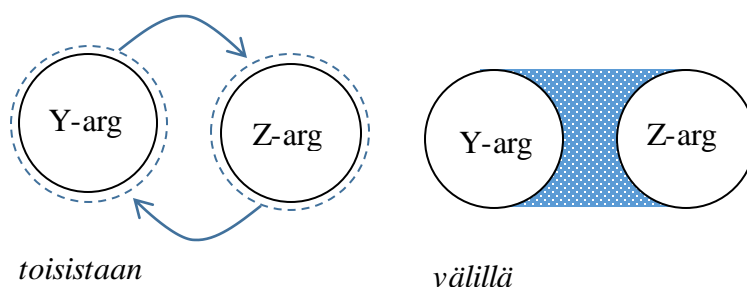
10e. Siitä vain **selvittämään, mikä on** vapaudenriiston ja kiinnioton välinen ero. (K1618)

*ero-*nominissa *kanssa* esiintyvistä resiprookkisista sanoista yleisin on *välillä* (10c), joka VISK:n (§ 689) mukaan on paikallissijamuotojen ilmaisumahdollisuuksia hyödyntävä adpositio. VISK puhuu sen lokatiivisesta käytöstä, jolloin se ilmoittaa suhteellista sijaintia, kuten lauseessa *Kiista johtajan ja alaisten välillä ~ kesken*. Aineistossani *välillä*-lekseemi esiintyy useassa sijamuodossa, myös tulosijassa (10d) ja muodossa *välille* (n=7).

Lisäksi aineistossani esiintyy näkökulmaistavia adpositioita (VISK § 689), sellaisia kuin *verrattuna* (n=5) ja *nähdessä* (n=4), jotka koodasin kategoriaan RESIPROOKKI, vaikka ne nähdäkseni hahmottavat pikemminkin ”yksisuuntaisen” kuin kahden kohteen välisen relaat-ion⁸². Kuviossa 3 olen hahmotellut lekseemien *toisistaan* ja *välillä* semantiikkaa.

⁸¹ *välinen*-sana on paikkaa tai aikaa ilmaiseva suhteutusadjektiivi (VISK:n § 613). *vs* on lyhenne latinan sanasta *versus*, jonka suomenkieliseksi vastineeksi Wikipedia (artikkeli ”Vs”) esittää lekseemin *vastaan*.

⁸² Ne esiintyvät rakenteissa (Y_arg.olosija+) *ero* + kopula + Z_arg.illatiivi, kuten esimerkissä *Maahan tullessaan rikollisella sen sijaan on vain yksi ero tavalliseen turistiin verrattuna: ajatukset* (K1530).



Kuvio 3. Lekseemien *toisistaan* ja *välillä* skemaattiset diagrammikuviot.

Kuviossa mustat ympyrät esittävät Y- ja Z-argumentteja. *välillä*-lekseemin kuviossa sininen alue vihjaa, että fokuksena on entiteettien välille jäävä, suuntaisuuden suhteen neutraali vyöhyke. *toisistaan*-lekseemin kuviossa siniset katkoviivaiset ympyrät ja niistä lähtevät nuolet vihjaavat, että fokuksena on vastavuoroinen suhde, jossa elatiivin ilmoittama yksisuuntainen separaatio ("Y lähtee pois Z:n luota"; *Y eroaa Z:sta*) ikään kuin kaksisuuntaistuu ("Y lähtee pois Z:n luota ja Z lähtee pois Y:n luota"; *Y ja Z eroavat toisistaan*). Yksisuuntaisessa separaatiossa Z-argumentti toimii LÄHDEVÄYLÄ:nä. Elatiivi spesifioi VÄYLÄ:n nimenomaan LÄHDEVÄYLÄ:ksi (havainnoinnin lähtöpisteeksi), ja täsmentää "liikkeen" suuntaisuuden (vrt. Leino 1993: 192). Vastavuoroiseksi separaatioksi kutsumassani käsitteistyksessä LÄHDEVÄYLÄ:n mielikuvaskaema sen sijaan hämärtyy. Jälkimmäisessä osallistujat esittävät VISK:n sanoja (§ 486) lainatakseni tasa-arvoisina, edellisessä yhdestä näkökulmasta.

4.3 Verbejä koskevat muuttujat

Koodauksen iso haaste on nominien ja verbien vertailu, jonka ongelmat korostuvat verbien ominaisuuksiin liittyvien muuttujien koodauksessa. Koska *ero*-sana on nomini, siltä uupuvat (*erota-* tai *erottaa-*) verbien muotokriteerit: se ei taivu tempuksessa, moduksessa tai persoonassa (Hakulinen & Karlsson 1979: 83). Noministatuksen vuoksi on keinotekoista, ellei kerettiläistä, koodata *ero*-nominille verbien ominaisuuksia mittaavien muuttujien arvoja.

Kielenkäytössä se on kuitenkin kytköksissä verbeihin, joiden muotokriteereitä on mahdollista mitata. Ensinnäkin *ero*-nomini toimii transitiiverbien objektina, jolloin verbimuuttujat mittaavat transitiiverbien ominaisuuksia. Toisekseen se esiintyy kopulalauseissa predikaatiivina tai subjektina, jolloin verbimuuttujat mittaavat kopula-verbin ominaisuuksia. Kopulalauseet ovatkin kiinnostavia. Vilkuna (1996: 90) kirjoittaa, että predikaatiivi "on sikäli erikoinen funktio, että se on semanttisesti katsoen itse lauseen predikaatti, ja lau-

seen subjekti sen (yleensä ainoa) argumentti. Predikatiivi on tässä mielessä analoginen intransitiiviselle verbille: vaikkapa ilmaus (*on*) *innokas* on semanttisesti samaa tyyppiä kuin ilmaus *hössöttää*.”. Voidaan pohtia, edustaako ilmaus (*on*) *eroa* semanttisesti ehkä samaa tyyppiä kuin ilmaus *erota*. Hypoteesini on, että edustaa. Demonstroidakseni *ero*-nominin ja *ero*-verbien välisen semanttisen analogian hypoteesia olen postuloinut aineistoni eli todellisen kielenkäytön mallin pohjalta joukon minimipareja tai -ketjuja taulukkoon 14.

Taulukko 14. *ero*-johdosten välisiä analogioita.

<i>ero</i> -nomini	<i>erota</i> -verbi	<i>erottaa</i> -verbi
Y:n ja Z:n ero on siinä, että ~ suuri Y:n ja Z:n välillä on suuri ero siinä, että Y:llä on se ero Z:hyn, että Y:ssä on eroa (siinä että).	Y ja Z eroavat suuresti ~ siinä, että Y eroaa Z:sta suuresti ~ siinä, että Y:yt eroavat toisistaan siinä että	Y:tä ja Z:aa erottaa se, että
Y:ssä ja Z:ssä ei ole mitään eroa	Y ja Z eivät eroa toisistaan mitenkään	
X ei tajua Y:n ja Z:n eroa . X ei tajua, mikä on Y:n ja Z:n ero		X ei osaa erottaa Y:tä Z:sta.
Jos [pariskunnalle] tulee ero .	Jos pariskunta eroaa .	
Monet vaativat Y:n eroa ministerin tehtävistä.	Y:n tulisi monen mielestä erota ministerin tehtävistä.	Y tulisi monen mielestä erottaa ministerin tehtävistä.

Silmäilemällä taulukkoa 14 havaitaan, että *ero*-johdoksilla todella on mahdollista predikoida samankaltaisia asiantiloja ja vieläpä samasta näkökulmasta, samat oliot näyttämöllä (vrt. Leino 1993: 80). Toisaalta *ero*-johdokset eroavat toisistaan perustavanlaatuisesti: on nomina, intransitiiviverbiä ja transitiiviverbiä, jotka poikkeavat toisistaan sanaluokkaominaisuuksiltaan ja syntaktiselta käyttäytymiseltään. Koodauksessa haaste onkin etsiä kultainen keskitie siinä, miten paljon informaatiota johdosten ”morfologisesta tai morfosyntaktisesta ontologiasta” lekseemienvälisiin (klusteri)analyysihin syötetään (ks. luku 8). Siksi työn painopiste ei ole *ero*-johdosten rakenteessa vaan niiden seuralaisissa. Tästä huolimatta liitin työhön joukon muotokriteerejä koskevia muuttujia (joskin rajasin monta ulos.)

Verbien muotokriteerejä mittaavat muuttujat ja niiden esittelyluvut ovat VERBI_RAKENNE (luku 4.3.1), PÄÄLUOKKA (luku 4.3.2), MODUS (luku 4.3.3), TEMPUS (luku 4.3.4) ja MODUS1 ja MODUS2 (luku 4.3.5). Tässä yhteydessä esittelen myös kieltoisanan läsnäoloa mittaavan muuttujan KIELTO (4.3.6).

4.3.1 VERBI_RAKENNE

Muuttuja VERBI_RAKENNE mittaa, millaisessa verbikonstruktiossa *ero*-johdos esiintyy. Muuttujan inspiraationa on Karlssonin yms. (1980) muuttuja (25) VERBIN RAKENNE, joskin

luokkajako on omatekoinen⁸³. Määritin muuttujalle neljä luokkaa, jotka ovat ERO.VERBI, VERBIKETJU, ON.VAIKEAA ja MUU. Taulukko 15 esittää luokkien jakauman lekseemeittäin.

Taulukko 15. Muuttujan VERBI_RAKENNE jakauma lekseemeittäin.

Muuttuja	Ala-arvot	\sum_{ero}	\sum_{erota}	$\sum_{erottaa}$	\sum_{pesye}
VERBI_RAKENNE	ERO.VERBI	522	397	257	1176
	VERBIKETJU	57	137	175	369
	ON.VAIKEAA	7	16	48	71
	MUU	8	13	29	50

Luokista yleisin on ERO.VERBI. Siihen koodasin korpuksimerkit, joiden finiittiverbinä toimii *erota-* tai *erottaa-*verbi yksin tai *ero-*nomini yksittäisen verbin yhteydessä. Määritin toisin sanoen luokan ERO.VERBI siten, että siihen koodautuvat lauseet sisältävät leinolais-langackerilaisittain ilmaistuna (ks. esim. Leino 1993: 72–75, 80–86, 111–112) kompositorakenteen, jonka rakenneskeema on muotoa $[[\text{ERO}]-[\text{'VERBILLISTIN'}]]$. Rakenneskeemassa predikaatin $[\text{ERO}]$ toteutumia ovat *ero-*kantanomini denominaalisissa johdoksissa *ero-TA* ja *ero-TTA* sekä nominissa *ero*. Kukin niistä sisältää osarakenteena (näin *erota-* ja *erottaa-*verbeissä) tai kokonaisrakenteena (näin *ero-*nominissa) atemporaalisen relaation $[\text{ERO}_1]$, joka vastaa juurimorfeemia *ero*, ja joka on niitä skemaattisemman predikaatin $[\text{ERO}]$ toteutuma. Predikaatin $[\text{'VERBILLISTIN'}]$ toteutumia ovat predikaatit $[\text{TA}]$ ja $[\text{TTA}]$, jotka spesifioivat verbijohdinten eli manifestoivat prosessuaalista predikaatiota. Predikaatin $[\text{'VERBILLISTIN'}]$ toteutumia ovat lisäksi kaikki skemaattisen predikaatin $[\text{VERBI}]$ toteumat eli yksiköt, jotka hahmottavat prosessuaalisen predikaatin – toisin sanoen sanat, joiden sanaluokkana on verbi. Rakenneskeeman $[[\text{ERO}]-[\text{'VERBILLISTIN'}]]$ toteutumia ovat rakenneskeemat $[[\text{ERO}_1]-[\text{VERBI}]]$, $[[\text{ERO}_1]-[\text{TA}]]$ ja $[[\text{ERO}_1]-[\text{TTA}]]$, joiden toteutumia taas ovat muun muassa rakenneskeemat $[[\text{ero}]-[\text{-tA}]]$, $[[\text{ero}]-[\text{-tA}]]$ ja $[[\text{ero}]-[\text{olla}]]$. Kooste tästä turhankin monimutkaiselta kuulostavasta erittelystä on esillä asetelmassa 1.

Asetelma 1. Luokan ERO.VERBI skemaattinen jäsenmys.

Skemaattisuus aste	juuri	verbijohdin tai verbi
Yläskeema	$[[\text{ERO}]]$	$[\text{'VERBILLISTIN'}]$
Väliskeemat (yläskeeman toteumat)	$[\text{ERO}_1]$	$[\text{TA}]$, $[\text{TTA}]$, $[\text{VERBI}]$
Kielellinen toteuma	<i>ero</i> , <i>ero-</i>	<i>-tA</i> , <i>-ttA</i> , <i>olla</i> , <i>ymmärtää</i>

Asetelmassa 1 kielellinen toteuma *ero* viittaa *ero-*nomiiniin, joka toimii sellaisten verbien kuin *olla* ja *ymmärtää* seuralaisena. *ero-* puolestaan viittaa *ero-*verbien kantasanaan, johon

⁸³ Karlssonin yms. (1980) alaluokat ovat: 1) yksi sana; 2) liittomuoto; 3) verbiketju; 4) verbiketju liittomuodossa; 5) *on aika* / *määrä* + inf., *katsoa hyväksi* + inf.; 6) edell. liittomuodossa; 7) fraasiverbi (*saada aikaan* yms.); 8) edell. liittomuodossa; 9) verbipari; 10) edell. liittomuodossa ja 11) finiittiverbin ellipsi.

erota-johdoksessa yhtyy verbijohdin -tA ja *erottaa*-verbissä -ttA. Skeeman [[ERO₁]-[TA]] spesifikaatioita ovat *erota*-verbin sisältävät lause-esimerkit, esimerkiksi lause *Lesbot eroaa siinä kuin heterot, mutta homot ei* (K982). Skeeman [[ERO₁]-[TTA]] spesifikaatioita ovat *erottaa*-verbin sisältävät lause-esimerkit, esimerkiksi lause *Se, mikä ihmisen erottaa eläimestä, on ajattelukyky* (K443). Skeeman [[ERO₁]-[VERBI]] spesifikaatioita ovat muun muassa seuraavat *ero*-nominin sisältävät lause-esimerkit:

11a. **On** sillä **ero**, teloitetaanko vajaat 1500 vaiko 10 000 plus vankileireillä kuolleet. (K1556)

11b. Sitten kun **ero tulee**, vaimo vie mieheltä puolet omaisuudesta. (K1658)

11c. **Tiedätkö**, mikä **ero on** sanoilla "moni" ja "kaikki"? (K1538)

Esimerkissä 11a on kyse kopulalauseesta, 11b:ssä intransitiivilauseesta ja 11c:ssä transitiivilauseesta eli rakenteista *ero* + kopula, *ero* + intransitiiviverbi ja transitiiviverbi + *ero* + kopula, jonka olen myös koodannut luokkaan ERO.VERBI. Muuttuja ei otakaan kantaa siihen, onko kyseessä intransitiivi-, transitiivi- vai kopulalause (tätä mittaa muuttuja LUOKKA). Muuttuja ei myöskään ota kantaa verbin liittomuotoihin. Siksi keksityt lauseet *Kissa ei ole eronnut koirasta* ja *Kissa eroaa koirasta* edustavat kumpikin luokkaa ERO.VERBI (vrt. Karlsson yms. [1980: 58] joka erottelee liittomuodot). Muuttuja sivuuttaa senkin, onko rakenneskeeman [[ERO]-['VERBILLISTIN']] toteuma leksikaalistuneessa (*erota*- ja *erottaa*-verbi) vai leksikaalistumattomassa asussa (*ero*-nominin ja verbin yhteenliittymä).

Toiseksi yleisin alakategoria on VERBIKETJU. Koodasin luokkaan rakenneskeeman [[[MODAALIVERBI]-[ERO]-['VERBILLISTIN']]] toteumat. Kyse on toisin sanoen *erota*-verbin, *erottaa*-verbin tai *ero*-nominin ja verbin sekä modaalisen apuverbin muodostamasta verbikonstruktiosta. Kun kyse on *erota*- tai *erottaa*-verbistä, ketjun loppuosana esiintyy A-infinitiivi tai MA-infinitiivi (ks. VISK § 496). Yleisimmät modaaliset apuverbit ovat *erottaa*-verbillä *pitää* (n=41), *voida* (n=40), *osata* (n=30), *tulla* (n=14); *erota*-verbillä *voida* (n=54), *pitää* (n=20) ja *tulla* (n=12); ja *ero*-nominilla *voida* (n=14), *pitää* (n=10), *tulla* (n=12) ja *osata* (n=7). Esimerkit 11d–11f havainnollistavat koodausratkaisuja:

11d. Nimikristittyjen **pitäisi erota** kirkosta, ei uskovaisten. (K914)

11e. Poliitikkojemme **kannattaisi opetella erottamaan** toisistaan sellaiset peruskäsitteet kuin pakko ja vapaus. (K508)

11f. Taksin ja 'valetaksin' **ero voi** pienimmillään **olla** vain se, että taksilla on toimilupa, valetaksilla ei. (K1572)

En ole tehnyt eroa siinä, koostuuko verbiketju kahdesta (11d) vai useammasta (11e) verbistä. *ero*-nominin yhteydessä apuverbi voi liittyä kopulaverbiin (11f) tai johonkin muuhun verbiin.

Luokassa ON.VAIKEAA on kyse verbikonstruktiosta, jossa *ero*-kantainen lekseemi täydennyksineen kytkeytyy *on ihanaa* ~ *vaikeaa* -rakenteeseen, kuten esimerkissä 11g:

11g. Rahanpesua on vaikea **erottaa** tavallisesta rahojen käyttämisestä. (K239)

Vallitseva lauseenjäsenäysperinne (ks. esim. VISK § 504) jäsentää *on vaikea* -rakenteeseen liittyvän infiniittisen lausekkeen infiniittisubjektiksi, koska niiden suhde verbiin on samanlainen kuin NP-subjektilla. Lauseenjäsenäys menee tällöin jotakuinkin näin: *Rahanpesua* [infiniittilausekkeen lohkeama] *on vaikea* [kopula + predikaatiivi] **erottaa tavallisesta rahojen käyttämisestä** [infiniittilausekkeen ydinosa]. Työssäni olen valinnut toisenlaisen jäsenäyksen. Katson rakenteen edustavan kiteytynyttä ilmaisutyyppiä, jossa *ero*-lekseemien täydennykset (tässä *rahanpesu* ja *tavallinen rahojen käyttäminen*) tulkittuvat Y- ja Z-argumenteiksi eli objektiksi ja rektioadverbiaaliksi. Ratkaisuani taustoittaa minimipari 11h. Minimipari 11i esittää tilanteen, jossa *on vaikeaa* -rakenne saa genetiivisubjektin (n=6)⁸⁴.

11h. *Rahanpesua*_{Y_ARG.} **on vaikeaa erottaa** *tavallisesta rahojen käyttämisestä*_{Z_ARG.} (K239) : *Rahanpesua*_{Y_ARG.} **ei voi erottaa** *tavallisesta rahojen käyttämisestä*_{Z_ARG.} (muokattu).

11i. *Suomalaisviranomaisten*_{X_ARG.} **olisi ollut vaikea erottaa** *Yhdysvaltain epäiltyjä vankilentoja*_{Y_ARG.} *muista lennoista*_{Z_ARG.} (K286) : *Suomalaisviranomaiset*_{X_ARG.} **eivät olisi osanneet erottaa** *Yhdysvaltain epäiltyjä vankilentoja*_{Y_ARG.} *muista lennoista*_{Z_ARG.} (muokattu)

Vaikka verbikonstruktiot ovat erilaisia, minimiparin 11h lauseista kumpikin on yleistävä: agenttia ei ole ja lauseet saavat ajasta ja paikasta irrallisen ”geneerisen” tulkinnan (vrt. Vilkuna 1980). Edellinen lause voidaan tulkita modaalisuudeltaan dynaamiseksi, kun taas muokatussa lauseessa *voida*-verbi luo ’on mahdotonta’-tulkintaa. Genetiivisubjekti minimiparissa 11i synnyttää analogian *on vaikeaa* -rakenteen ja perussubjektin sisältävien rakenteiden välille, jota havainnollistaa minimiparin jälkimmäinen jakso (muokattu lause).

Luokka MUU on ylijäämäluokka, joka sisältää muun muassa rakenteet, joissa *ero*-lauseke toimii infiniittisubjektina (11j) tai nominilausekkeen laajenuksena (11k) (ks. VISK § 563).

11j. Muualla maailmassa on käytäntönä tosiaankin [**erottaa** vääryyteen tai rikokseen syyllistynyt oppilas.]_{infiniittisubjekti} (K31)

11k. Pitäisi palauttaa presidentin oikeus [**erottaa** työstään hallitus ja PM.]_{infiniittilauseke NP:n laajenuksena} (K364)

Vallitsevan lauseenjäsenäysradition näkökulmasta luokat ON.VAIKEAA ja MUU sisältävät kumpikin lauserakenteita, joissa *ero*-lauseke muodostaa infiniittisubjektin. Tällä perusteella luokat on yhdistetty monimuuttujaisessa analyysissä.

⁸⁴ Paikallissijaista tekijäkonstituenttia esiintyy niin ikään 6 kertaa.

4.3.2 PÄÄLUOKKA

Muuttuja PÄÄLUOKKA tarkastelee, onko lauseen finiittiverbi tai finiittiverbi, jonka vaikutusalassa *ero*-lauseke esiintyy, aktiivi- vai passiivimuotoinen. Luokkajako ja lekseemikohmainen jakauma on esillä taulukossa 16.

Taulukko 16. Muuttujan PÄÄLUOKKA luokkajako ja jakauma lekseemeittäin.

Muuttuja	ID-tasot	Σ_{ero}	Σ_{erota}	$\Sigma_{erottaa}$	Σ_{pesye}
PÄÄLUOKKA	AKTIIVI	560	547	395	1502
	PASSIIVI	34	16	114	164

On ilmeistä, että kielessä aktiivimuoto on odotuksenmukainen. Myös aineistossani passiivi edustaa ”poikkeustilaa”. Koodausratkaisujani havainnollistavat esimerkit 12a–12d:

12a. Brittiekonomistin neuvo: Näin eurosta erotaan käytännössä. (K621)

12b. Henkilö, joka vahingoittaa yhdistyksen mainetta, voidaan erottaa yhdistyksestä. (K189)

12c. Lähes 38000 ihmistä oli allekirjoittanut adressin, jossa vaaditaan Räsäsen **eroa** ministerin tehtävästä. (K1242)

12d. Ei osata tehdä eroa työmatkakuluihin ja työajokorvauksiin jne. (K1073)

Olen koodannut muuttujan arvoksi PASSIIVI, kun *erota*- tai *erottaa*-verbi esiintyy yksin yksipersonaisessa passiivissa (12a) tai passiivissa olevan verbiketjun jälkiosassa (12b). *ero*-nominin arvoksi olen koodannut PASSIIVI, kun *ero*-nomini esiintyy yksipersonaisessa passiivissa olevan verbin (12c) tai verbiketjun (12d) objektina tai subjektina.

4.3.3 MODUS

Muuttuja MODUS tarkastelee sen verbilausekkeen modusta, jonka vaikutusalaan *ero*-lauseke kuuluu. Taulukko 17 esittää muuttujan luokkajaon ja jakauman.

Taulukko 17. Muuttujan MODUS alaluokat ja jakauma lekseemeittäin.

Muuttuja	ID-tasot	Σ_{ero}	Σ_{erota}	$\Sigma_{erottaa}$	Σ_{pesye}
MODUS	INDIKATIIVI	556	476	426	1458
	MUU	38	87	83	208

Koodasin oman luokan ainoastaan INDIKATIIVI:lle, joka on kielessä yleisin modus. Luokka MUU on yhdistelmäluokka, joka kattaa konditionaalin, potentiaalin ja imperatiivin, joista kahdella viimeisellä on vain joitakin esiintymiä koko aineistossa. Koodausratkaisuja havainnollistavat esimerkit 13a–13c.

13a. Sukupuolella ei pitäisi olla eroa väkivallan tuomittavuuden suhteen. (K1204)

13b. Voisiko työpaikastakin **erota** ilmoittamalla jollekin välityssivustolle. (K615)

13c. Muutto erottaisi meidät vanhemmistani, jotka ovat ainoat lastenkaitsijamme. (K466)

Niin kuin esimerkistä 13a näkyy, lauseen koodausarvo määräytyy sen verbin mukaan, jonka täydennyksenä *ero*-lauseke on. Jos verbilausekkeeseen kuuluu useampi verbi (13a–13b), joista jokin on esimerkiksi konditionaalissa, arvoksi koodautuu MUU. Jos verbejä on vain yksi, arvo määräytyy tämän verbin moduksen mukaan (13c).

4.3.4 TEMPUS

Muuttuja TEMPUS osoittaa muuttujan MODUS tavoin sen verbin aikamuoto, jonka vaikutusalassa *ero*-lauseke on. Päädyin jakamat tempukset kolmeen luokkaan, jotka ovat PREESENS, PERFEKTI ja MUU (imperfekti tai pluskvamperfekti), koska pluskvamperfektillä on aineistossani hyvin vähän esiintymiä ja koska pluskvamperfektin skeemakohtainen jakauma muistuttaa imperfektin jakaumaa. Skeemoissa, joissa imperfekti on harvinainen, ei esiinny pluskvamperfektiäkään. Taulukko 18 esittää muuttujan TEMPUS jakauman lekseemeittäin.

Taulukko 18. Muuttujan TEMPUS alaluokat ja jakauma lekseemeittäin.

Muuttuja	ID-tasot	Σ_{ero}	Σ_{erota}	$\Sigma_{erottaa}$	Σ_{pesye}
TEMPUS	PREESENS	522	444	424	1390
	MUU	48	79	50	177
	PERFEKTI	24	40	35	99

4.3.5 MODUS1 ja MODUS2

Modaalisuutta mittaavia muuttujia ovat morfologista verbimodusta tarkasteleva muuttuja MODUS (luku 4.3.3) sekä MODUS1 ja MODUS2, jotka tarkastelevat modaaliverbejä ja modaalisen tulkinnan saavia rakenteita (nesessiivirakenne) sekä modaalisia adjektiveja (*mahdollinen, välttämätön*) – sivuuttaen partikkelit (*ehkä, luultavasti*). En käytä jakoa dynaamiseen, deonttiseen ja episteemiseen modaalisuuteen (ks. VISK § 1551–1614). Sen sijaan modaalisuuden teoreettisena pohjana työssäni toimii kaikista yksinkertaisin dikotomia välttämättömyyden (*necessity*) ja mahdollisuuden (*possibility*) välillä (ks. Arppe 2008: 341; Kangasniemi 1992: 6; VISK § 1551). Taulukko 19 esittää muuttujien jakauman.

Taulukko 19. Muuttujien MODUS1 ja MODUS2 lekseemikohtainen jakauma.

Muuttuja	ID-tasot	Σ_{ero}	Σ_{erota}	$\Sigma_{erottaa}$	Σ_{pesye}
MODUS1	VÄLTÄMÄTTÖMYYS	36	51	88	175
	MAHDOLLISUUS	25	58	52	136
	NA	533	454	369	1355
MODUS2	KYKY	45	7	110	162
	NA	549	556	399	1504

MODUS1 sisältää kaksi luokkaa, VÄLTÄMÄTTÖMYYS ja MAHDOLLISUUS, ja MODUS2 yhden luokan, KYKY (*ability*), jonka Arppe (mts. 342) sisällyttää kategorian MAHDOLLISUUS piiriin.

Havaitaan, että luokkien VÄLTÄMÄTTÖMYYS ja MAHDOLLISUUS havainnot jakautuvat tasaisesti kuin luokan KYKY, joka keskittyy lekseemeihin *ero* (n=45) ja *erottaa* (n=110)⁸⁵. Taulukosta 20 käy ilmi muuttujien yleisimmät spesifikaatiot eli rakennetyypit.

Taulukko 20. Muuttujien MODUS1 ja MODUS2⁸⁶ yleisimmät rakennetyypit lekseemeittäin.

	<i>ero</i>	<i>erota</i>	<i>erottaa</i>
VÄLTÄMÄTTÖMYYS	<i>vaatia</i> (n=17), <i>pitää</i> (n=8), <i>kannattaa</i> (n=5)	<i>pitää</i> (n=20), <i>tulee</i> (n=12), <i>tarvita</i> (n=2)	<i>pitää</i> (n=41), <i>tulee</i> (n=13), <i>kannattaa</i> (n=6)
MAHDOLLISUUS	<i>voida</i> (n=16), <i>taitaa</i> (n=4), <i>mahtaa</i> (n=2)	<i>voida</i> (n=53)	<i>voida</i> (n=41), <i>on mahdollista ~ mahdotonta</i> (n=4).
KYKY	<i>ymmärtää</i> (n=27), <i>osata tehdä</i> (n=7), <i>pystyä</i> (n=5), <i>tajuta</i> (n=4)	<i>tajuta</i> (n=3), <i>ymmärtää</i> (n=2), <i>osata</i> (n=1)	<i>osata</i> (n=32), <i>olla vaikea[a]</i> (n=25), <i>kyetä ~ on kyettävä ~ olla kyky ~ kykenemätön</i> (n=21)

Taulukosta havaitaan, että olen aika mekaanisesti tulkinnut joitain verbejä ja rakenteita modaalisiksi. Prototyyppisesti verbit *osata*, *kyetä* ja *pystyä* signaloivat kykyä, *voida*-verbi ja *on mahdollista* -rakenne mahdollisuutta ja verbit *pitää*, *tulee* ja *täytyy* välttämättömyyttä. Koska modaaliverbit kuitenkin ovat polyseemisiä, koodausratkaisut on arvioitu tapauskohtaisesti kieli-intuition nojalla. (Malli)esimerkkejä luokasta KYKY ovat 14a–14c, joista 14a sisältää *erottaa*-verbin, 14b *erota*-verbin ja 14c *ero*-nominin.

14a. Jos todella olet lekuri, **etkä osaa erottaa** vanhuutta sairaudesta, niin [...]. (K332)

14b. Jos tämä mies **ei ymmärrä erota** virastaan, niin on Suomi uskomaton paikka. (K1002)

14c. Armeija **ei osaa tehdä eroa** simputuksen, koulutuksen, sotilaskurin välillä. (K1213)

⁸⁵ Modaalilla aineksilla puhuja ilmaisee, onko asiantila hänen mielestään, epävarma, välttämätön, (epätodennäköinen), mahdollinen, pakollinen, luvallinen, ulkoisista tai sisäisistä edellytyksistä riippuvainen ja niin edelleen (VISK § 1551). Työssäni modaalisuutta tarkastellaan suppeasti. Esimerkiksi Arppe (2008: 341–347) jakaa luokan POSSIBILITY positiiviseen mahdollisuuteen (PROPOSSIBILITY; *voida, olla mahdollisuus*) ja mahdotto-muuteen (IMPOSSIBILITY; *ei voi, ei ole mahdollisuutta*). Lisäksi Arppe erottaa luokat luvalle (PERMISSION; *saada, olla lupa*) ja suosituksella (PROHIBITION; *ei saa, olla kiellettyä*). Luokka NECESSITY sisältää jaon positiiviseen välttämättömyyteen (PRONECESSITY; *pitää, täytyä, joutua*) ja negatiiviseen (NONNECESSITY; *ei tarvitse*) sekä hyödyttömyyteen (FULITY; *ei kannata, on turha*). Lisäksi hänellä on luokat tahdonalaisuudelle (VOLITION; *haluta, yrittää*) ja temporaalisuudella (TEMPORAL; *ehtii, olla aika[a]*).

⁸⁶ Muuttuja MODUS2 on oman muuttujanaan eikä muuttujan MODUS1 alaluokkana koodausteknisistä syistä – siksi, että modaalisuuden ilmaisukeinot ketjuuntuvat usein tavalla, jota VISK (§ 1560) nimittää modaaliharmoniaksi. Modaaliharmoniaa demonstroivat esimerkit 14A–C.

14A. **Pitää**_{VÄLTÄMÄTTÖMYYS} vain **osata**_{KYKY} **erottaa** murhanhimo ja murha toisistaan. (K212)

14B. Tämä **pitäisi**_{VÄLTÄMÄTTÖMYYS} **voida**_{MAHDOLLISUUS} todentaa, ja **erottaa** [...] pysäköinnistä. (K230)

14C. Meidän **pitää**_{VÄLTÄMÄTTÖMYYS} **pystyä**_{KYKY} **tekemään ero** kahden pysäköintitapahtuman välillä. (K1436)

Esimerkeissä 14A ja 14C ketjuuntuvat VÄLTÄMÄTTÖMYYS ja KYKY, esimerkissä 14B VÄLTÄMÄTTÖMYYS ja MAHDOLLISUUS. Koska en luonut yhdistelmäluokkaa VÄLTÄMÄTTÖMYYS + MAHDOLLISUUS ja koska muuttujasysteemi rajoittaa modaalikombinaatiot yhdistelmiin VÄLTÄMÄTTÖMYYS + KYKY (n=18) ja MAHDOLLISUUS + KYKY (n=1), olen 14B:n kaltaisissa tapauksissa koodannut modaalisuuden verbiketjun ensimmäisen jäsenen perusteella. Modaaliharmonia esiintyy lisäksi nesessiivirakenteen ja lekseemin yhteisvaikutuksesta, kuten esimerkeissä Näissäkin asioissa **on kyettävä eroittamaan** ne osatekijät, jotka [...] (K469) ja Katsojan **on** aina **voitava erottaa** mainokset toimituksellisesta sisätekijöistä (K325). Edeltävä kytkeytyy luokkiin KYKY ja VÄLTÄMÄTTÖMYYS, jälkimmäinen saa arvon luokassa MAHDOLLISUUS.

Esimerkeissä 14a ja 14c modaalitulkinnan synnyttää *osata*-verbi. Esimerkki 14b edustaa yhtä *erota*-verbin seitsemästä havainnosta, joissa verbiketjussa kuuluu aina jokin seuraavista verbeistä: *tajuta* (n=3), *ymmärtää* (n=2), *osata* (n=1) tai *älytä* (n=1). *ero*-nominin koodaus poikkeaa siinä mielessä yleisestä linjasta, että olen uhmakkaasti koodannut luokkaan KYKY myös tapaukset, joissa lauseen finiittiverbinä esiintyy *ymmärtää*-verbi (n=20) tai *tajuta*-verbi (n=4) joko paljaassa lauseessa (14d) tai yhdyslauseessa (14e):

14d. Jos **et tajua eroa** lastenkirjan ja miljoonien harjoittaman elämäntapavalinnan välillä, niin [...] (K1132)

14e. **Onko** kirjoittaja ymmärtänyt, **mikä ero on** pedofiililla ja opettajalla joka kiusaa ja lanistaa oppilastaan. (K1523)

Jos verbiin kytkeytyy kyky-tulkintainen modaalinen adverbis (n=5), kuten lauseessa *Venäläisiä ei erota (helposti) ulkonäön perusteella*. (K113), tai predikaatiivadjektiivi, kuten esimerkissä (K111) *Kiusaamista on myös vaikea erottaa silloin, kun kiusaaja ja kiusattu liikkuu samoissa piireissä*, esiintymä on koodattu luokkaan KYKY.

Luokan VÄLTÄMÄTTÖMYYS malliesimerkkejä demonstroivat esimerkit 14f–14j, joista 14f sisältää *erottaa*-verbin, 14g *erota*-verbin ja 14h *ero*-nominin.

14f. Kaikki neljä johtajaa pitäisi kyllä **erottaa** tämän sotkun seurauksena. (K36)

14g. Suomen tulee erota EU:sta ja liittyä sitten kun asiat ovat kunnossa. (K1044)

14h. Kun esitetään hyvä ajatus, osa vaatii esittäjän **eroa**. (K1284)

Esimerkeissä 14f ja 14g modaaliluennan laukaisevat *tulee*-verbi ja *pitää*-verbi, jonka konditionaal⁸⁷ merkitystä en ole eritellyt. *ero*-nominin koodaus poikkeaa muista siinä, että olen koodannut luokkaan KYKY lauseet (14h), joiden finiittiverbinä esiintyy *vaatia*-verbi (n=17). Olen siis tulkinnut, että *vaatia*-verbin semantiikkaan sisältyy nesessiivinen elementti. Luokkaan olen lisäksi koodannut havainnot, kuten esimerkin *Soinin kannattaisi erottaa Hakkarainen puolueesta* (K346), joissa *tarvita*- tai *kannattaa*-verbi ilmoittaa praktista välttämättömyyttä tai toimintasuositusta sekä havainnot, kuten esimerkin *Uupumuksen ehkäisemisen ja korjaamisen kannalta on tärkeää erottaa uupumuksen syyt* (K44), joissa adjektiivipredikaatiivin sisältävä *on hyvä ~ syytä ~ tärkeää ~ tarpeen* -rakenne ilmoittaa modaalista välttömyyttä⁸⁸. Luokan MAHDOLLISUUS (malli)esimerkkejä ovat seuraavat:

14k. Kelan valtuutettuja **ei voi erottaa** kansanedustajan luottamusvirasta. (K37)

14m. Lapsi voi erota kirkosta aikuisena, jos haluaa. (K577)

⁸⁷ Koko aineistossani luokan VÄLTÄMÄTTÖMYYS esiintymistä 43 % yhdessäesiintyy luokan MODUS_MUU kanssa (joka tarkoittaa käytännössä konditionaalia).

⁸⁸ Arppella (2008: 343) luokka HYÖDYTTÖMYYS (FUTILITY) kattaa spesifikaatiot *ei kannata* ja *on turha*.

14o. Kyllä tyttöjen ja poikien välillä **voi olla** suurin **ero** yläkoulussa. (K1594)

Näissä *voida*-verbi laukaisee mahdo(llisuuteen/-ttomuuteen) liittyvän modaalisen tulkinnan. Muistettakoon, että modaalisuutta mittaavat muuttujat alaluokkineen (kuten työni kaikki luokat) ovat prototyyppisiä kategorioita, jotka projisoivat jakaumaa, jonka keskiössä ovat mallijäsenet, reunoilla perifeeriset tapaukset. Usein on epäselvää, kuuluuko esiintymä yhteen vai toiseen luokkaan tai ehkä luokkien välille vedenjakajaksi, tapaukseksi, jolla on eri luokkien ominaisuuksia ja joka voidaan tulkita usealla tapaa⁸⁹ (ks. Raukko 1994: 53).

4.3.6 KIELTO

Muuttuja KIELTO mittaa lauseen kielteisyyttä. Taulukko 21 esittää muuttujan luokkajaon lekseemikohtaisen jakauman sisällyttäen *ero*-nominin molemmat sijamuodot.

Taulukko 21. Muuttujan KIELTO jakauma ja luokkajako.

Muuttuja	ID-tasot	Σ_{ero}	Σ_{eroa}	Σ_{erota}	$\Sigma_{erottaa}$	Σ_{pesye}
KIELTO	NEUTRAALI	346	93	456	374	1269
	KIELTEINEN	19	136	107	135	397

Taulukkoa 21 silmäilemällä havaitaan, että muuttuja KIELTEINEN on jonkin verran yleisempi *erottaa*-verbin kuin *erota*-verbin yhteydessä (kytkeytyen erityisesti skeemaan EROTTAA.HAVAITA2). Lisäksi havaitaan, että nimenomaan *eroa*-muoto on kytkeytynyt kielteisiin lauseisiin. Tämä johtuu siitä, että sekä objektin että predikaatiivin sija kielteisissä lauseissa on partitiivi. Esimerkit 15a ja 15b havainnollistavat koodausratkaisuja:

15a. Jos joku (esim. poliisi?) **ei tiedä, mitä eroa** noilla kahdella **on**, niin [...]. (K1260)

15b. **Ei** internet **eroa** muusta mediasta niin paljon kuin kuvitellaan. (K647)

Esimerkin 15a tavoin muuttuja saa arvokseen KIELTEINEN, mikäli finiittiverbilauseke, jonka vaikutusalueella *ero*-lauseke on, sisältää kieltoisanan. Kun lause sisältää *erota*- tai *erottaa*-verbin, kuten esimerkki 15b, arvo määräytyy verbin kielteisyyden pohjalta.

Siirryn seuraavaksi tarkastelemaan adverbiaaleja koskevia muuttujia.

4.4 Adverbiaaleja koskevat muuttujat

Jo kandidaatin tutkielmassani havaitsin, että *erota*- ja *erottaa*-verbien kanssa esiintyy paljon tapaa, aikaa ja keinoa ilmaisevia adverbiaaleja. Sisällytin niitä työhöni siitäkin huolimatta,

⁸⁹ Esimerkin 14B (ks. s. 75, alav. 86) minimipariksi voidaan postuloida useampiakin lauseita, kuten *tämä pitäisi pystyä jotenkin todentamaan ja erottamaan* tai *tämä pitäisi olla mahdollista jotenkin todentaa ja erottaa*. Esimerkin *voida*-verbin olisi siis ehkä voinut koodata myös muuttujan MODUS2 luokkaan KYKY.

että osalla niistä on niukasti esiintymiä ja vinoutunut jakauma. Esimerkiksi muuttuja SYY kytkeytyy lähes yksinomaan skeemoihin, jotka ilmaisevat (työ-)suhteen tai romanttisen suhteen päättymistä. Sivuutan ne yksimuuttujaisesta vaiheesta, mutta en monimuuttujaisesta vaiheesta, sillä klusterianalyysi ei aseta tiukkoja ehtoja luokkien koolle (vrt. kuitenkin Glynn 2012; Gries 2006). Taulukko 22 esittää adverbiaaleja mittaavat muuttuja ja niiden jakauman.

Taulukko 22. Adverbiaaleja mittaavat muuttujat SYY, AIKA, KVANTTORI ja VASTAUS. NA-arvot ovat poissa luettavuuden takia.

Muuttuja	ID-tasot	<i>ero</i>	<i>erota</i>	<i>erottaa</i>	\sum_{pesye}
SYY	LÄSNÄ	17	46	30	93
	NA				(1573)
AIKA	LÄSNÄ	24	99	69	185
	NA				(1474)
KVANTTORI	LÄSNÄ	245	90	9	334
	NA				(1322)
VASTAUS	LÄSNÄ	110	33	31	174
	PREDIKOIVA	49	-	-	49
	NA				(1443)

Aloitan tarkastelemalla muuttujaa SYY (luku 4.4.1), josta siirryn muuttujan AIKA (4.4.2) kautta muuttujiin KVANTTORI (luku 4.4.3) ja VASTAUS (luku 4.4.4).⁹⁰

4.4.1 SYY

Muuttuja SYY mittaa tilanteen syytä. Yleisimmät rakennetyypit ovat *johdosta* (n=2), *jälkeen* (n=1), *seurauksena* (n=4), *syystä* (n=7), *takia* (n=19), *vuoksi* (n=13), *perusteella* (n=2), *siksi* (n=5) ja *kun ~ koska* (n=37). VISK:n (§ 1128: asetelma 161) kausaalista suhdetta (eli syyn ja seurauksen merkitystä) ilmaisevien merkitsimien luettelossa on listattuna edellisistä muut paitsi *jälkeen* ja *perusteella* sekä temporaaliset lauseenvastikkeet (n=3), jotka koodasin ryhmään *kun ~ koska*. Esimerkit 16a–16c hahmottavat koodausratkaisuja:

16a. Kummallista mikä haloo nousee, jos ihminen ilmaisee mielipiteensä jostain asiasta ja sen takia ollaan nyt **vaatimassa** Räsäsen **eroa**. (K1262)

16b. Kehitysvammaisten lasten vanhemmat **eroavat** useammin kuin muut johtuen varmaankin siitä, että sairaan lapsen kanssa elämä on vain liian raskasta. (K576)

16c. Thaimaalais-suomalaisissa perheissä perheväkivalta on yleistä, **eivätkä** naiset **uskalla erota**, koska pelkäävät menettävänsä oleskelulupansa, varsinkin jos heillä on lapsia. (K777)

Esimerkissä 16a muuttujan SYY spesifikaationa toimii *sen takia* -lauseke, joka on (*sen*) *vuoksi* -lausekkeen rinnalla yleisimpiä spesifikaatiota. (Ne edustavat nähdäkseni kategorian

⁹⁰ Työni laajuuden vuoksi rajasin analyyseistä ulos monia adverbiaaleja mittaavia muuttujia, esimerkiksi *eroa*-misen keinoa tai välinettä tai reittiä ilmaisevat muuttujat KEINO, VÄLINE ja REITTI sekä muuttujan VASTAUS semanttisena sukulaisena toimivan muuttujan KOHDE, joka esiintyy lauseessa *Kissalla ja koiralla on eroa karvanlaadussa*. Mahdollisessa jatkotutkimuksessa niiden läsnäolo olisi tietysti toivottavaa (ks. luku 8).

prototyyppiä.) Esimerkissä 16b SYY:tä edustaa *johtuen siitä* -lauseke. Ryhmiin *siksi ja kun* ~ *koska* kuuluvien lausekkeiden koodauksessa (16c) rikoin periaatettani sivuuttaa analyysistä lauserajan ylittävät ainekset

Koska olen koodannut kategoriaan SYY ison joukon erilaisia kielenaineiksia ja koska useita niistä (esimerkiksi *koska*-sanaa) voidaan käyttää monenlaisissa tehtävissä, on SYY-kategoria luonnollisesti heterogeeninen. Ilmeisesti kategoria on kuitenkin tarpeeksi yhtenäinen. En esittele SYY:tä tämän enempää, vaan siirryn KVANTTORI:n tarkasteluun.

4.4.2 KVANTTORI

Muuttuja KVANTTORI (n=350) on dikotominen kategoria. Positiivisen koodausarvon eli LÄSNÄ (n=350) saavat *erota*-verbin yhteydessä esimerkit (n=87), jotka sisältävät eroavaisuuden paljoutta tai toiminnan intensiteettiä spesifioivan adverbiaalilausekkeen. Yleisimpiä rakennetyyppejä ovat: *mitenkään* (n=14), *juuri(kaan)* (n=12), *paljon* (n=9), *kuin yö (ja päivä/päivästä)* (n=4), *niin paljon (että)* (n=5), *suuresti* (n=3), *täysin* (n=3). *ero*-nominin yhteydessä (n=242) muuttujan yleisimmät spesifikaatiot ovat: *mitään* (n=34), *suuri* (n=27), *merkittävä* (n=19), *ainoa* ~ *ainut* (n=18), *pieni* (n=16), *vissi* (n=11), *iso* (n=9), *selvä* (n=8), *kuin yöllä ja päivällä* (7)⁹¹. *erottaa*-verbillä KVANTTORI:a esiintyy yhdeksästi. *erota*-verbin koodausesimerkkejä hahmottavat esimerkit 17a ja 17b:

17a. Työ tutkimuslaitoksen ja yliopiston välillä **eroaa** kuin yö päivästä. (K678)

17b. En usko, että Vekarajärvi **eroaa** mitenkään muista varuskunnista. (K645)

Esimerkissä 17a KVANTTORI:n spesifioi *kuin*-vertailurakenne ja esimerkissä 17b *mitenkään*-adverbiaali. Tällaisissa lauseissa KVANTTORI ilmaisee erilaisuuden intensiteettiä

⁹¹ Näihin sekakoosteisiin ryhmiin sisältyy joukko kvanttiadiverbeja ja partikkeleita, esimerkiksi määrän suuruutta ilmaisevia adverbeja (*paljon, hieman, tarpeeksi*), kieltohakuisia partikkeleita (*ollenkaan, lainkaan, mitään*), täyteen asteen adverbiaaleja (*kokonaan, täysin, tyystin*), adjektiivistuneita partisiippeja (*merkittävä, huomattava, ratkaiseva*) sekä lisäksi *kuin*-rakenteellisia vertailuilmauksia (*kuin yö päivästä*), suhteellisia adjektiiveja (*pieni, iso, suuri*), modaalisia adjektiiveja (*selvä, ilmeinen*), komparatiivadjektiiveja (*isommin*) sekä numeraalin sisältäviä lausekkeita (*yksi merkittävä, parikin*) ja yksilöiviä adjektiiveja (*ainoa* ~ *ainut, tietty*), jotka rajaavat tarkoitteen potentiaalisten tarkoitteiden joukosta. Joukossa esiintyy myös *niin x että* -rakenteellisia konsekutiivia konstruktia (*niin pirusti, että*), jossa alisteinen *että*-lause täsmentää ominaisuuden astetta mm. seurauksen kannalta, sekä saman vertailuasteen ilmaisemiin käytettyjä konstruktia *niin x kuin* (*niin paljon kuin kuvitellaan*). (ks. tarkemmin VISK § 565, 583–584, 605–608, 632–635, 657–664, 789, 819, 1160, 1166, 1173; erityisesti asetelmat 102 ja 108). En ole eritellyt vertailuasteita (*suuri* : *suurempi* : *suurin*), intensiteettimääreitä (ks. VISK § 664, asetelma 100), kuten insentiteettipartikkeleita (*aika, melko, varsin*) ja geneetiivimuotoja (*turhan pieni, hyvin selkeä*), enkä sitä, onko lausekkeen edussanana adjektiivi (*suuri*), adverbii (*suuresti*) tai pronomini (*mitään*). En ole koodannut minkään muuttujan alle irrallisia kommenttiadverbiaaleja, kuten asenneadverbiaaleja (*onneksi, toivottavasti, valitettavasti*), tai episteemistä modalisuutta ilmaisevia adverbeja ja partikkeleita (*epäilemättä, ehkä, ilmeisesti, tietenkin, varmaankin*), enkä irrallisia konnektiiveja (*toisin, silti, kuitenkin, siitä huolimatta*). (ks. VISK § 605, 667, 676, 963.)

”kirjaimellisesti” (17b) tai metaforisesti (17a). Vain kolmesti muuttuja esiintyy muun skeeman kuin EROTA.IRTI yhteydessä. Niitä ovat skeemaan EROTA.IRTI2 lause *Sitten kun kirkko eroaa valtiosta kokonaan* (K634), skeeman EROTA.ABSTRAKTI virke *Te-toimisto ja Kela kulkevat käsikädessä, joten erosin näiden lafkojen piiristä täysin* (K665) sekä skeeman EROTA.KIRKKO lause *Lisäksi kirkosta erotaan niin paljon [...] (K649)*. Edelliset kaksi ilmoittavat toiminnan täydellisyyden astetta, viimeinen kvantifioi määrää eli kirkosta eroajien paljoutta. Jälkimmäistä tyyppiä, johon kuuluvat myös lausekkeet *entistä enemmän* ja *koko ajan lisää* (n=3), esiintyy vain *erota*-verbillä. Merkitys on eri kuin lauseessa *VR:n toiminta eroaa tosi paljon oluen valmistuksesta* (K666), jossa *paljon* ilmoittaa erilaisuuden intensiteettiä.

ero-nominin koodausesimerkkejä hahmottavat esimerkit 17c–17e:

17c. Yhteisöllisyydellä ja anarkialla **on vissi ero**. (K1631)

17d. Suomen ja Kreikan **ero** ei sittenkään **ole suurensuuri**. (K1564)

17e. Ihmisessä **ei ole mitään eroa**⁹² muiden nisäkkäiden kanssa, paitsi väärän toimintaa taipuvat aivot. (K1074)

ero-nominin yhteydessä KVANTTORI:n spesifikaatio on syntaktiselta rooliltaan määrite⁹³ (n=193), kuten esimerkeissä 17c ja 17e tai predikatiivi (n=50), kuten esimerkissä 17d. *ero*-nominin ja *erota*-verbin välillä ilmenee systemaattista vaihtelua, jota havainnollistavat parit *mitään : mitenkään* ja *suuri : suuresti*. Tyyppi *ainoa ~ ainut* (n=18) – joka esiintyy esimerkissä *Ainoa ero muihin uskontoihin on* (K1331) – luo assosiaatioita *erottaa*-verbiin, jossa tyyppiä esiintyy kaksi kertaa, kuten esimerkissä *Ainut mikä hänet erottaa tavallisesta lääkäristä on se, että tutkinto puuttuu* (K268). Assosiaatiota *erota*-verbin skeemojen EROTA.IRTI2 ja EROTA.ABSTRAKTI syntyy suhteessa täyteen asteen adverbiaaleihin (n=3). Esimerkissä [*Yhdysvalloissa*] *valtio on täysin erotettu kirkosta* (K128) täyteen astetta spesifioi *tässä*-sana.

On syytä kysyä, mittaako muuttuja samaa ilmiötä kussakin lekseemissä tai edes yhden lekseemin sisällä. Ilmeistä on, että kyse on semanttisesti ja syntaktisesti heterogeenisestä kategoriasta, jonka piiriin kuulu monenlaisia määrää ja intensiteettiä modifioivia sanoja. Aluksi muuttujalla olikin useita alaluokkia, joista osa – vaikkapa TÄYTEYDENASTE (n=5) –

⁹² *eroa*-muodon kanssa KVANTTORI (n=55) esiintyy lauseenloppuisena (n=5) ainoastaan vertailuilmauksissa, jotka ovat tyyppiä *tsekkioluissa eroa on kuin yöllä ja päivällä* (K1129). Määritteenä esiintyvistä *eroa*-muodon KVANTTORI-tapauksista (n=50) 90 % sisältää kieltoisan ja 94 % esiintyy *olla*-verbin jälkeen loppukentässä, kuten esimerkissä *naisella ja miehellä ei ole yhtikäs mitään eroa esim. pappina toimimiseen* (K1200) (ks. VISK § 1370); 66 % spesifikaationa *on mitään*-lekseemi. *ero*-muodon KVANTTORI-esiintymistä (n=142) 70 % sijaitsee *olla*-verbin jälkeen loppukentässä, alle 4 % sisältää kieltoisan; *mitään*-sanaa ei esiinny kertaakaan.

⁹³ Koodasin esiintymän määritteeksi, kun se sijaitsee *ero*-nominin edessä riippumatta siitä, missä lauseenjäsenyspaikalla *ero*-sana on. Predikatiiviksi koodasin esiintymät, jotka sijaitsevat reemapaikalla – *ero*-nominin ja kopula-verbin välillä – tai näiden välissä, kuten lauseen *Ei tuo ero dramaattinen ole* (K1234).

sai hyvin vähän esiintymää. Määrällisen analyysityökalujen rajoitusten vuoksi päädyin kärkeään dikotomiseen luokitteluun. Nähdäkseni KVANTTORI mittaa kussakin lekseemissä ”tarpeeksi samanlaista” semanttista ilmiötä sopien saman liukumaluonteisen kategorian sisälle.

4.4.3 VASTAUS

Muuttuja VASTAUS mittaa, esiintyykö tarkastellussa aineistoesimerkissä lauseke, joka ilmoittaa kahden entiteetin välisen yhtäläisyyden tai eron. Kyse on senkaltaisesta propositiosta, joka on ajateltavissa vastaukseksi kysymykseen, mitä eroa on kissalla ja koiralla. ”Prototyyppiä” edustaa rakenne *se ~ siitä ~ siinä*-tukipronomini + *että*-lause.

erottaa-verbin (n=31) yhteydessä yleisimmät rakennetyypit ovat: *se, että*⁹⁴ (n=8), (*infiniitti*)*lauseke* (n=6), *joka_mikä* (n=4) ja *se* (n=3). Esimerkit 18a–18b havainnollistavat rakennetyyppejä:

18a. Työn **erottaa** harrastuksesta se, että siitä saa palkkaa. (K274)

18b. Nimenomaan kyky ajatella **erottaa** meidät eläimistä. (K448)

Esimerkissä 18a muuttujan VASTAUS spesifioi X-argumenttina toimia *se*-tukipronomini + *että* -rakenne, 18b:ssä taas *infiniittilauseke* *nimenomaan kyky ajatella*. Erikoista näissä koodausratkaisussa on se, että kielinelementit saavat arvon sekä X-argumenttia koskevissa muuttujissa (X_SEM ja X_MUOTO) että VASTAUS-muuttujassa. Ne koodaavat siis samanaikaisesti useampien muuttujien arvoja.

erota-verbissä muuttujan VASTAUS (n=33) yleisimmät rakennetyypit ovat: *siinä että*⁹⁵ (n=12), ”välimerkki” (n=3), *siinä* (n=2), *näin* (n=2), *tässä* (n=2) ja *siten että* (n=2). Esimerkit 18c–18d havainnollistavat koodausratkaisuja:

18c. Väinämöisen manifestit **eroavat** Breivikin kirjoituksista siinä, että niissä väkivallan kohteena nähdään ennen kaikkea maahanmuuttajat._{VASTAUS} (K861)

18d. Tilastointi eri maissa **eroaa**, esim. Etelä-Euroopassa ei virallisten rikostilastojen mukaan ole ollenkaan kotiväkivaltaa._{VASTAUS} (K846)

Esimerkissä 18c muuttujan VASTAUS spesifioima kielinelementti esiintyy adverbiaali(täydennyksenä) ja esimerkissä 18d itsenäisessä lauseessa, joka on välittömästi *ero*-lekseemillisen lauseen perässä.

⁹⁴ VISK:n (§ 1145) mukaan tukipronomini (yleensä neutraali *se*) korjaa tietyissä lauseasemissa sitä puutetta, että lauseeseen ei voi suoraan liittää sijapäätettä. Tällöin sijapäätteen saa yksikkömuotoinen pronomini, joka toimii lausekkeen edussanana ja jota alisteinen lause määrittää. VISK:n mukaan tukipronominillisella ja paljaalla lauseella ei aina ole selvää merkityseroa. Myös tutkimuksessani tukipronomini edustaa pikemminkin vakiintunutta käytännettä, enkä oletta sen kantavan isompaa itsenäistä merkitystä.

⁹⁵ VISK:n (§ 1156) mukaan ”tukipronominin tehtävissä toimivat myös sellaiset pronominin sisältävät ilmaukset kuin *siinä, sen suhteen, siihen nähden* ja *siltä kannalta*; [aineistossani lisäksi *siinä mielessä, siinä suhteessa, sillä tavalla, siltä osin*]; pronomini ei kuitenkaan ole välttämätön”.

ero-nominin yhteydessä muuttujan VASTAUS (n=110) spesifioi suurin valikoima erilaisia rakennetyyppejä, joista yleisimmät ovat: *se että* (n=31), *siinä että* (n=17), ”kaksoispiste” (n=11), *että* (n=6), *se* (n=6), *tämä ~ tässä* (n=6) sekä proadjektiivin sisältävä konsekutiivinen konstruktio, *sellainen ~ semmoinen, että* (n=6), jossa *että*-lauseen ilmaisema asiantila esitetään subjektitarkoitteen (*ero*-nominin) ominaisuutena (ks. VISK:n § 725, 1161). Esimerkit 18e–18j havainnollistavat VASTAUS-muuttujan koodausratkaisuja.

18e. Pieni mutta oleellinen **ero on** myös, se että hampaat [...] ja hörökorvat ovat jatkuvasti näkyvissä/muistuttamassa itseään ja muita. (K1311)

18f. Hoitajien ja lääkärien kiireessä **on se ero**, että lääkärin kiire on aikataulutettu. (K1422)

18g. **Ero** kuoleman ja köyhyyden välillä **on siinä, että ihminen ei kuoleman jälkeen joudu kamppailemaan ongelmien kanssa vaikka se vie mahdollisuuden nauttia elämästä. (K1452)**

18h. Sekä ympärileikkaus että lapsen lyöminen tuottavat lapselle kipua. **Ero on**, että ympärileikkaus jättää pysyvän vamman. (K1383)

18i. Aloissa **on** kuitenkin **eroa**: taidehistorioitsijat työllistyvät heikommin. (K1111)

18j. Heidän politiikassaan **ole juuri muuta eroa** muihin kuin maahanmuuttokriittisyys. (K1103)

Esimerkki 18e muuttujan VASTAUS spesifikaationa toimii *se, että* -rakenne, 18f:ssä lohjennut *se, että* -rakenne, 18g:ssä adverbiaali(täydennys) (joka koostuu tukipronominista *siinä* ja alisteisesta *että*-lauseesta), 18h:ssä paljas eli tukipronominin *että*-lause ja 18i:ssä kaksoispisteellä erotettu itsenäinen lause. Esimerkissä 18j muuttujan VASTAUS spesifioi *muuta x kuin* -lauseke.

On syytä todeta, että muuttujan VASTAUS havainnot jakautuvat sillä tavalla epätasaisesti eri sijamuotoihin, että kopulalauseessa (VASTAUS, n=99) *ero*-muodolla on 92 esiintymää, kun taas *eroa*-muodolla on 7 esiintymää. Lisäksi *eroa*-muodossa esiintyy vain rakennetyyppejä *ei muuta kuin (että)* (18j) ja *sen verran, että* (joita ei esiinny *ero*-muodon kanssa) sekä rakennetyyppejä ”kaksoispiste”⁹⁶. Muuttuja VASTAUS assosioituukin *ero*-muotoon.

Lopuksi on syytä pohtia, mittaako muuttuja VASTAUS yhtä ja samaa ilmiötä kussakin lekseemissä, kuten toivottavaa on. Sekä *erottaa*-verbillä että *ero*-nominilla esiintyy tyyppejä *se että, että* ja *lauseke*, joista ensin mainittu ja yleisin koodautuu *erottaa*-verbillä subjektiksi, *ero*-nominilla predikatiiviksi, joskin sen jäsentäminen subjekti olisi myös mahdollista. *erota*-verbin yleisintä tyyppiä *siinä että* sekä tyyppiä *tämä* esiintyy myös *ero*-nominilla.

⁹⁶ Tyypit *sen verran eroa, että* ja *muuta eroa kuin (että)* (18j) assosioituvat muuttujan KVANTTORI tyyppeihin *niin paljon ~ suuri, että* (vrt. luku 4.4.3). Kummassakin on *ero*-lekseemin etumäärirteenä toimiva elementti, jota täydentää alisteinen *että*-lauseen. Olen koodannut luokan VASTAUS esiintymiksi kuitenkin vain edelliset, sillä jälkimmäisissä alisteiset *että*-lauseet edustavat seurausta, selitystä yms.

Vaikka lauseenjäsennysroolit vaihtelevat, ainakin samojen (tuki)pronominien varaan rakentuvat lausekkeet lienevät semanttiselta funktioltaan läheisiä. Jäljelle jää kysymys, vastaavatko esimerkiksi *erottaa*-verbin tyyppi *se että* vaikkapa *erota*-verbin tyyppettä *siinä mielessä, että* tai *tässä*. Ei täydellisesti, mutta riittävästi kuuluakseen samaan kategoriaan.

4.4.4 AIKA JA YLEISET MUUTTUJAT

Koodasin dikotomisen kategorian AIKA luokaksi LÄSNÄ silloin, kun lauseessa esiintyy adverbiaali, jonka merkitykseen nimenomaisessa kontekstissa sisältyy niin sanoakseni temporaalinen hahmo (vrt. Leino 1993: 224). Temporaaalisia hahmoja kuvastavat 19a–19e:

- 19a. 80-luvulla pa-sairaiden vanhemmista **erosi** 80 %. (K511)
- 19b. Kun kirkosta luultavasti lähivuosina **erotaan** entistä enemmän. (K580)
- 19c. Helmikuun puolivälistä lähtien **on ollut** ennenäkemätön **ero** säissä. (K1524)
- 19d. Muualla maailmassa poliitikko[...] **erotettaisiin välittömästi**. (K90)
- 19e. Kirkosta saa omalla päätöksellä **erota** 18-vuotiaana. (K697)

Esimerkissä 19a kyse on rajatusta ajanjaksosta (*torstaina, tammikuussa, vuonna 1809, viikon aikana*), 19b:ssä epämääräisestä ajanjaksosta (*nykyään, aikoinaan, joskus, tällä hetkellä*), esimerkissä 19c tapahtuman alku- tai päätepisteestä (*lopulta; lapsesta saakka*), esimerkissä 19d tapahtumahetkestä (*nyt, heti*) ja esimerkissä 19e iästä. (ks. VISK § 649–652, 689, 980; erityisesti asetelmat 106, 113 ja 151.)⁹⁷AIKA spesifioi yhden kategorian piiriin paitsi suhteellisen erilaisia aikasuhteita myös erilaisia rakenteita, kuten olosijaisia adverbiaaleja (ks. VISK § 1248) ja essiivisijaisia predikatiivadverbiaaleja (ks. VISK § 1258). Iso osa lausekkeista on puiteadverbiaaleja eli lauseen integroituja osia, koska ne sisältyvät mm. kiellon ja kysymyksen vaikutusalaan (ks. VISK § 962). Sellaiset puiteajanmääreet edustavatkin luokan prototyyppiä (eniten esiintymiä). Perifeerisempiä jäseniä (joille aluksi koodasin omat luokat) ovat OSMA:t (ks. VISK § 972, 448; Leino 1991: 51) sekä 19fn kaltaiset tapaukset, jossa kvantifoidaan agenttia:

- 19f. Jos olen **eronnut** kolme kertaa ulkomailla, olenko Suomessa bigamisti? (K575)
- 19g. Eduskuntaryhmä **erotti** hänet pariksi viikoksi. (K423)

Iteratiivisia ajanmääreitä (n=8), jotka ilmoittavat toiminnan toistoisuutta (19f), esiintyy vain *erota*-verbillä. 19g:n kaltaisia aikaa rajaavia OSMI:a esiintyy *erottaa*-verbillä (n=6)

⁹⁷ Jätin koodaamatta ’joissakin tapauksissa’ -tyyppiset ajanilmaukset, joista useat edustavat kvantifioivia proadverbeja (ks. VISK § 741). Koodaamatta jäivät muun muassa esimerkin *Sen ja ihmiskaupan ero voi joskus olla vähäinen* (K1542) sana *joskus* ja esimerkin *Kommenteista näyttää pudonneen pois joitakin muotoiluja, lainausta ei aina erota* vastauksesta sana *aina* (K96).

ja *erota*-verbillä (n=2, viikon aikana ja 23-vuoteen). Heterogeenisyydestään huolimatta AIKA on nähdäkseni tarpeeksi yhtenäinen muuttuja edustaakseen yhtä ja samaa kategoriaa.

Lopuksi tarkastelen vielä niin sanottuja yleisiä muuttujia, joita ovat LUOKKA, KORPUS ja LUKU. Ne ja niiden jakauma ovat esillä taulukossa 23.

Taulukko 23. Muuttujat LUOKKA, KORPUS ja LUKU ja niiden jakauma.

Muuttujat	ID-tasot	Σ_{ero}	Σ_{eroa}	Σ_{erota}	$\Sigma_{erottaa}$	Σ_{pesye}
LUOKKA	TRANSITIIVI	54	151	0	509	713
	INTRANSITIIVI	46	5	563	0	615
	KOPULA	265	73	0	0	338
LUKU	K1–K1666	365	229	594	509	1666
KORPUS	KOMMENTTI (353 eri kommentoijaa)	–	–	–	471	–
	HS.FI	–	–	–	38	–

Muuttujista LUKU toimii tunnisteena, joka identifioi tilastoyksiköt K-arvoilla. Muuttujan KORPUS, joka mittaa kirjoittajaa, koodasin vain *erottaa*-verbille. Kirjoittajat ovat jakautuneet siten, että eri nimimerkkejä, jotka ovat jättäneet yhden kommentin, on yhteensä 309, kaksi kirjoitusta jättäneitä on 40 nimimerkillä, kolme 14 nimimerkillä, neljä 8 nimimerkillä, kuusi yhdellä nimimerkillä ja 38 yhdellä nimimerkillä, joka on HS.fi. Esitutkimuksessa havaitsin, ettei merkitysvarianttien vaihtelussa ilmene merkittävää roolia sillä, onko tilastoyksikkö peräisin toimituksen tuottamasta uutisesta vai kommentoivasta tekstiosasta. Muuttuja LUOKKA mittaa, minkälaisessa lausetyypissä *ero*-johdos esiintyy.

Luvussa 4 esiteltujen muuttujien lisäksi työssäni joitakin tukimuuttujia, jotka esittelen tapauskohtaisesti. Siirryn nyt yksimuuttujaisen analyysiin.

5 Yksimuuttujainen analyysi ja sen tulokset

Aloitan yksimuuttujaisen analyysin esittelemällä X^2 -riippumattomuustestin, jota luvussa 5.2 sovellan *erota*-verbin, luvussa 5.3 *erottaa*-verbin ja luvussa 5.4 *ero*-nominin analyysiin. Analyysit ovat sillä tavalla identtisiä, että analysoin kunkin lekseemin kohdalla X-, Y- Z-argumentteja koskevat muuttujat sekä muuttujat RAKENNE ja RESIPROOKKI. Työni laajuus ei mahdollista muiden muuttujien yksimuuttujaista analyysia. *ero*-lekseemin analyysissä on mukana tukimuuttujia, joita ei esiinny *ero*-verbeillä. Analyysien lopussa esittämässäni yhteenvedoissa sivuan lisäksi muuttujia, kuten PÄÄLUOKKA, jota analysoin tarkemmin luvussa 6. Luvussa 5.5 kokoan ajatukset yksimuuttujaisen analyysivaiheen tuloksista.

5.1 X^2 -riippumattomuustesti

Yleisin korpuslingvistiikassa hyödynnettävistä yksimuuttujaisista testeistä lienee X^2 -riippumattomuustesti (*chi square test of independence*, jatkossa *khiin neliö -testi* tai *X^2 -testi*). Sillä testataan, poikkeako havaittu jakauma satunnaisesta jakaumasta (odotetuista frekvensseistä) (ks. Gries 2015c: 9). Khiin neliö -testin ideana on muodostaa kaksi hypoteesia eli mahdollista tulosta – nollahypoteesi (H_0) ja vaihtoehtoinen hypoteesi (H_1). Nollahypoteesi muodostetaan lähtöhypoteesin vastaisesti olettamalla kaiken muuttujien arvoissa havaitun vaihtelun johtuvan sattumasta. Testauksen tarkoituksena on selvittää, kumpi hypoteeseista saa tukea aineistosta eli jää voimaan. Hypoteesi ei koskaan jää voimaan absoluuttisesti, vaan se jää voimaan jollain todennäköisyydellä (p - tai α -arvo). Ihmistieteissä yleinen kriittinen todennäköisyys eli merkitsevyystaso on $\alpha=0.05$ ⁹⁸. α -arvo ilmoittaa, kuinka suurella todennäköisyydellä vaihtoehtoinen hypoteesi on väärä. Työssäni käytetty $\alpha=0.05$ tarkoittaa viiden prosentin erehtymistodennäköisyyttä, jolloin toistettaessa testiä 20 kertaa on todennäköistä, että kerran kahdestakymmenestä nollahypoteesi hylätään perusteetta (mikä edustaa virhetyyppiä I)⁹⁹. (Luoja 2006: 87–88; Nummenmaa 2004: 136–138.) X^2 -testissä muodostetut hypoteesit ovat (ks. Pitkänen & Kohonen 1984: 156):

H_0 . Tarkastellut muuttujat ovat toisistaan riippumattomia ($\alpha \leq 0.05$), toisin sanoen havaitut ja odotetut frekvenssit eivät poikkea toisistaan merkitsevästi; ei ole havaittu eroja.

H_1 . Tarkastellut muuttujat ovat toisistaan riippuvia, toisin sanoen havaitut ja odotetut frekvenssit poikkeavat toisistaan merkitsevästi ($\alpha > 0.05$); on havaittu eroja.

Testissä pienten erojen voidaan katsoa johtuvan satunnaisvaihtelusta, suurten taas jostakin systemaattisesta tekijästä (Pitkänen & Kohonen mts. 154). Odotettujen ja havaittujen frekvenssien lisäksi testiin tarvitaan tieto vapausasteista (*df*, *degree of freedom*), joka saadaan kaavalla (kontingenssitaulukon rivien lukumäärä-1)*(sarakkeiden lukumäärä-1). 4x4-taulukossa se tarkoittaa laskua: $(4-1)*(4-1) = 9$. (ks. Nummenmaa 2004: 291.)

⁹⁸ p -arvoja määritellään toisinaan seuraavasti: $p < 0.001$ = 'erittäin merkitsevä', $0.001 \leq p < 0.01$ 'hyvin merkitsevä', $0.01 \leq p < 0.05$ merkitsevä', $0.05 \leq p < 0.1$ 'marginaalisesti merkitsevä' (ks. Gries 2013: 29).

⁹⁹ Tyypin I virheen lisäksi on olemassa tyypin II virhe, joka tarkoittaa tilannetta, jossa vaihtoehtoinen hypoteesi H_1 hylätään, vaikka sen olisi pitänyt jäädä voimaan. Toisin sanoen ilmiö esiintyy populaatiossa mutta se jää huomaamatta. Näistä virheistä tyypin I pidetään yleisesti vaarallisempana. (ks. Nummenmaa 2004: 140.)

X^2 -testin oletuksiin kuuluu, että kyseessä on populaatiosta poimittu sattumanvarainen otos¹⁰⁰. Kaikkien muuttujien täytyy lisäksi olla nominaaliasteikollisia ja kaikkien odotettujen frekvenssien pitää olla suurempia kuin 1 ja ainoastaan 20 prosentissa luokista saa olla ≥ 5 ¹⁰¹. (Nummenmaa 2004: 290; Cochran 1954: 420.) Jos käyttöoletukset tai ”minimiehdot” eivät täyty, tuloksiin tulee suhtautua varauksella (Heikkilä 2008: 213). Pitkänen ja Kohonen (1984: 157–158) esittävät kolme keinoa tilanteeseen, jossa minimiehdot eivät täyty. Tutkija voi i) yhdistellä luokkia suuremmiksi, ii) osittaa taulukon ja suorittaa testin siinä osassa jossa se on luvallista (mikäli testaaminen on yhä mielekästä) tai iii) vaihtaa tilastollista testiä.

Ennen analyysiin ryhtymistä on syytä muistaa Pitkäsen ja Kohosen (ks. 1984: 153, 167) sanat siitä, että kielentutkimuksessa kielen kuvaus on pääasia. Tilastollisen analyysin rooli on palvella kuvausta ja tehostaa argumentointi. Karlsson yms. (1980: 94) taas muistuttaa, etteivät tilastollinen ja lingvistinen merkitsevyys välttämättä lankea yksin: mikä on tilastollisesti merkitsevää ei aina ole lingvistisesti merkitsevää [tai kiinnostavaa], ja mikä on lingvistisesti merkitsevää ei aina noudata ankaria tilastollisia merkitsevyysvaatimuksia.

5.2 erota-verbi

Khiin neliö -testin vuoksi *erota*-verbin skeemoja on yhdisteltävä¹⁰² niin, että mukana ovat skeemat: EROTA.ERI, EROTA.IRTI (\leftarrow EROTA.IRTI2, $n=8$ ¹⁰³), EROTA.KIRKKO ja EROTA.PUOLISO. Mikroskeemat EROTA.ABSTRAKTI ($n=3$) ja EROTA.KONKRETIA ($n=2$) eivät ole mukana analyysistä, mutten usko sen vinouttavan tuloksia. Tässä onkin käytetty edellisen luvun keinoista i) luokkien yhdistelyä ja ii) taulukon osittamista (vrt. Karlsson yms. 1980: 93–96).

¹⁰⁰ Kuten Gries (2015b: 101) muistuttaa, korpuslingvistiikassa käytetyt menetelmät kuten X^2 -testaus ja logistinen regressio rikkovat tilastotieteen fundamentaaleja lähtöoletuksia, nimittäin sitä, että populaatio olisi täysin itsenäistä ja näytteet toisistaan riippumattomia. Ensinnäkin korpuksessa on yleensä yhdeltä kirjoittajalta/puhujalta enemmän kuin yksi tietue, mikä on totta myös aineistossani. Toiseksi tietyn lekseemin grammaattinen rakenne aiheuttaa idiosynkraattista käyttäytymistä, joka toistuu korpuksessa. Kolmanneksi korpuksessa on usein siinä merkityksessä hierarkkisesti rakentuneita, että eri kirjoittajat tai puhujat on klusteroitu eritasoisiksi rekistereiksi. Baayen (2001: 7, 162–167) nostaa erityisesti suurten korpuksien ongelmaksi lisäksi sen, että yksittäiset sanat ovat jakautuneet teksteihin epätasaisesti (johtuen nimenomaan tekstien koheesiosta ja diskurssirakenteesta, ei syntaktista). Kilgarriff (2005: 266) taas huomauttaa, ettei korpus ole sanojen sattumanvarainen sanavarasto, joka manifestoi khiin neliön matemaattista distribuutiota; kielenilmiöiden distribuutio noudattaa pikemminkin *Zipfian*: suurissa korpuksissa usein esiintyviä sanoja on vähän ja harvoin esiintyviä sanoja on paljon. Kilgarriffin mukaan khiin neliön (ja siihen liittyvän p-arvon) mittaaminen korpusaineistosta on yritys, johon testiä ei ole suunniteltu. Luojola (2006: 178) puoltaa X^2 -testiä ja mainitsee sen soveltuvan myös ”pienemmille korpuksille”, koska siinä ei oleteta sanataajuuksien jakautuvan normaalisti.

¹⁰¹ Arppe (2008: 78 alv. 47) tulkitsee minirajoituksia siten, että 8-solua sisältävässä 2x4-tilastossa yksi solu per muuttuja ($80\% \cdot 8 = 1.6$ solua) voi saada arvon ~ 1 silloin, kun loput kolme saa arvon ≥ 5 .

¹⁰² Kategorioiden yhdistely on intuitiivista, muttei ”mutuillua”. Koska tarkoitus on tehdä semanttisesti mielekkäitä luokitusratkaisuja, yhdistelyjen pohjana ovat aina havainnot kategorioiden välisistä suhteista.

¹⁰³ Notaatio KATEGORIA1 (\leftarrow KATEGORIA2) tarkoittaa, että KATEGORIA 2 on yhdistetty KATEGORIA:an 1.

Analyysi etenee siten, että luvussa 5.2.1 tarkastelen *erota*-skeemojen ja muuttujien Y_SEM ja Y_MUOTO välistä assosiaatiota. Luvussa 5.2.2 siirrän tarkastelun muuttujiin Z_SEM ja Z_MUOTO. Luvussa 5.2.3 tarkastelen muuttujia RAKENNE ja RESIPROOKKI, jotka tutkivat *erota*-verbin argumenttirakennetta. Luvussa 5.2.4 teen yhteenvedon tuloksista.

5.2.1 Muuttujat Y_SEM ja Y_MUOTO

χ^2 -testausta varten muokkasin muuttujan Y_SEM alakategoriat siten, että poistin luokan NA (n=104), jolloin jäljelle jäivät kategoriat ELOTON, KONKRETIA, INSTITUUTIO, JOUKKO ja YKSIÖ. Poistin NA-luokan – kuten toimin jatkossakin – koska se on täydellisessä korrelaatiossa muuttujan Y_MUOTO luokkien NOLLA ja NA kanssa. Tilastoyksikköjen määrä tippui toimenpiteiden takia 563:sta 454:een, joten analyysi ei kata koko *erota*-aineistoa. Taulukko 24 esittää luokkien jakauman *erota*-skeemoissa. Koska absoluuttisten esiintymistajuuksien avulla on hankala vertailla skeemojen välisiä suhteita, taulukossa 24b ilmenee (tummalli taustalla) kunkin alakategorian prosentuaalinen jakauma skeemojen sisällä (siten, että n=454 = 100 %) ja luokkien prosentuaalinen jakauma skeemojen välillä (vaalealla taustalla)¹⁰⁴.

Taulukko 24. Muuttujan Y_SEM havaintojen jakauma sekä harmaalla pohjalla prosentuaalinen jakauma skeemoittain (n=454). Isommat arvot on lihavoitu

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO	\sum_m
ELOTON	136	0	0	0	136
KONKRETIA	23	0	0	0	23
INSTITUUTIO	18	32	0	0	45
JOUKKO	18	16	34	37	105
YKSIÖ	7	80	44	14	145
\sum_s	197	128	78	51	454

Taulukko 24b. Muuttujan Y_SEM prosentuaalinen jakauma skeemojen välillä (vaalea tausta) ja skeemojen sisällä (tumma tausta).

	ERI		IRTI		KIRKKO		PUOLISO		\sum_m
ELOTON	100.0	69.0	0.0	0.0	0.0	0.0	0.0	0.0	30.0 %
KONKRETIA	100.0	11.7	0.0	0.0	0.0	0.0	0.0	0.0	5.1 %
INSTITUUTIO	28.9	6.6	71.1	25.0	0.0	0.0	0.0	0.0	9.9 %
JOUKKO	17.1	9.1	15.2	12.5	32.4	43.6	35.2	72.5	23.1 %
YKSIÖ	4.8	3.6	55.2	62.5	30.3	56.4	9.7	27.5	31.9 %
\sum_s	42.4 %		28.2 %		17.2 %		11.2 %		100.0 %

Silmäilemällä¹⁰⁵ taulukkoja 24 ja 24b havaitaan, että luokkien esiintymät eivät jakaudu tasaisesti. Ilmeisin havainto on, että luokkia ELOTON ja KONKRETIA esiintyy vain skeemassa

¹⁰⁴ Skeemojen välistä jakaumaa luetaan vaakasuunnassa ja skeemojen sisäistä pystysuunnassa. 24b:ssä rivi \sum_s esittää skeemojen osuuden ja sarake \sum_m luokkien osuuden kokonaisjakaumasta.

¹⁰⁵ Luojola (2006: 35) muistuttaa, etteivät aineistoa silmä määräisesti erittelevät menetelmät, kuten sanallinen päättely, ole tilastollista päättelyä. Sanallisten keinojen avulla voidaan kuitenkin saada yleiskuva aineistosta.

EROTA.ERI. Luokan INSTITUUTIO havainnot jakautuvat skeemoihin EROTA.ERI (28.9 %) ja EROTA.IRTI (71.1 %) ¹⁰⁶. Taulukosta 24b havaitaan, että myös skeemojen sisällä instituutiolla on suurempi osuus skeemassa EROTA.IRTI (25.0 %) kuin skeemassa EROTA.ERI (6.6 %). Taulukosta 24b havaitaan myös, että luokkaan YKSIÖ koodattuja agentteja suosivat erityisesti skeemat EROTA.IRTI (61.6 %) ja EROTA.KIRKKO (57.1 %), luokkaan JOUKKO koodattuja agentteja taas skeemat EROTA.KIRKKO (43.6 %) ja EROTA.PUOLISO (72.5 %). Luokkien EROTA.IRTI ja EROTA.KIRKKO välille syntyy eroa suhteessa luokkiin INSTITUUTIO – jota ei esiinny skeemassa EROTA.KIRKKO – ja JOUKKO, jota edustaa 13.0 % EROTA.IRTI skeeman agenteista. Näiden lisäksi havaitaan, että skeema EROTA.ERI kattaa yksinään lähes puolet (42.4 %) skeemojen kokonaispopulaatiosta. Luokista yleisimpiä ovat YKSIÖ (31.9 %) ja ELOTON (30.0 %). Khiin neliö -testin käyttämät odotetut arvot ovat esillä taulukossa 24c.

Taulukko 24c. Muuttujan Y_SEM odotetut arvot.

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO
ELOTON	59.0	38.3	23.3	15.3
KONKRETIA	9.9	6.5	3.9	2.6
INSTITUUTIO	19.5	12.7	7.7	5.0
JOUKKO	45.3	29.6	18.0	11.8
YKSIÖ	62.9	40.8	24.9	16.3

Havaittujen ja odotettujen frekvenssien jakaumaa silmäilemällä huomataan esimerkiksi, että luokan YKSIÖ havaintoja on skeemassa EROTA.ERI noin 56^{107} odotettua vähemmän ja skeemassa EROTA.IRTI noin 40 odotettua enemmän. Tulokseksi X^2 -testissä saadaan: $X^2=444.774$, $df=12$, $p < 2.2e^{-16}$. Tulkinta on, että taulukon 24c solujen X^2 -jakaumien summa 433.7045^{108} on selvästi suurempi kuin khiin neliön kriittinen arvo, $X^2 (df=12 \text{ ja } \alpha=0.05) = 21.026^{109}$. Lisäksi $p (X^2=433.7045, df=12) = 2.2e^{-16}$ (0.000000000000000022), joka on selvästi työn kriittistä merkitsevyystasoa ($\alpha = 0.05$) pienempi. Tulos antaa syyn hylätä nollahypoteesin (kriittisyystasolla 0.05) ja päätellä, että 1) taulukossa ilmenevä jakauma ei todennäköisesti

¹⁰⁶ Luokkien prosenttijakaumia voidaan tarkastella skeemojen välillä tai skeeman sisäisesti. Skeemojen väliset absoluuttiset ja prosentuaaliset vertailut eivät ole suoraan vertailukelpoisia, sillä tilastointiyksikköjen määrät skeemoissa vaihtelevat (taulukossa 24b 51:sta 197:een). Tuloksia raportoidessa on huomioitava, kuinka suuresta populaatiosta prosentiosuudet on laskettu. Isompi populaatio on luotettavampi ja edustavampi kuin pieni.

¹⁰⁷ Odotetut arvot – jotka lasketaan kaavalla: (rivin summa * sarakkeen summa) / esiintymien summa – saadaan automaattisesti moderneista tilastointiohjelmista, kuten R:stä komennolla *chisq.test(taulukko)\$expected*. Tekstissä esitetty erotus (56) on saatu laskusta: odotettuarvo (62.9) miinus havaittu arvo (7) = 62.9-7=55.9

¹⁰⁸ Laskettu X^2 -testisuure on merkitsevä, vaikka riskitasoa nostettaisiin $\alpha=0.001$, jolloin $df=12$, $X^2=32.909$. Pitkänen ja Kohonen (1984: 154) kuitenkin painottavat, että riskitaso on lyötävä lukkoon ennen tutkimusta.

¹⁰⁹ X^2 -testaus löytyy vakiona kaikissa moderneissa tilastointiohjelmista. R:ssä komento on *chisq.test*. Kriittiset merkitsevyystasot löytyvät jokaisesta tilastotiedettä sivuavasta oppaasta. Olen poiminut ne R:stä.

ole syntynyt sattumalta, että 2) homogeenisyyden sijaan olemme havainneet heterogeenisyyttä ja että 3) muuttujat ja skeemat eivät ole toisistaan riippumattomia. Sen sijaan niiden välillä on assosiatiivisuutta tai kytköksisyyttä.

Yksi X^2 -testin ongelmista on, että se kertoo vain, että jotain merkittävää on löydetty, muttei sitä, missä assosiaatio sijaitsee, minkä suuntaista ja miten vahvaa se on (Arppe 2008: 80–82). Sijainnin ja suunnan selvittämiseksi Arppe (mts. 81–84) esittää kolme strategiaa. Ensinnäkin voidaan tutkia, miten paljon yksittäisten solujen X^2 -jakaumat poikkeavat odotetuista arvoista. On olemassa konservatiivinen ja epäkonservatiivinen lähestymistapa. Konservatiivinen tapa on tutkia, ylittävätkö jotkin yksittäiset solut khiin neliön kriittisen merkitsevyystason samoilla vapausasteilla (df) ja p -arvolla (α). Epäkonservatiivinen tapa on tarkastella jokaista solua ikään kuin omana taulunaan, jonka vapausasteitten määrä df on 1, mikä tarkoittaa pienempää khiin neliön kriittistä merkitsevyystasoa (ja siten enemmän merkitseviä löytöjä). Kolmas keino on laskea standardoidut Pearsonin jäännökset (*standardized Pearson residuals*). Kyse on merkitsevistä poikkeamista (*a significant deviation*), kun jäännös on vähintään 2 absoluuttisesti (eli +2 tai -2) tai 3, jos kyseessä on ”iso taulukko” (Agresti 2002: 81). Kussakin strategiassa on mahdollista tutkia solukohtaisen merkitsevyyden lisäksi assosiatiivisuuden suuntaa, eli ylittävätkö vai alittavatko tarkasteltu arvo kohdearvon. Havainnollistan strategioita taulukolla 24f. Taulukko 24d esittää solukohtaiset X^2 -kertymät, jota konservatiivinen ja epäkonservatiivinen tarkastelu hyödyntää laskeessaan, ylittyykö kynnystaso vai ei. Standardoidut Pearsonin jäännökset hyödyntää taulukon 24e arvoja.

Taulukko 24d. Taulukon 24 Standardoidut Pearsonin jäännökset.

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO
ELOTON	+15.8	-8.6	-6.4	-4.9
KONKRETIA	+5.4	-2.9	-2.2	-1.7
INSTITUUTIO	-2.1	+6.9	-3.3	-2.5
JOUKKO	-6.2	-3.2	+4.6	+8.9
YKSIÖ	-11.0	+8.3	+5.3	-0.9

Taulukko 24e. Taulukon 24 Solukohtainen X^2 -kertymä.

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO
ELOTON	+98.5	-37.1	-23.5	-15.1
KONKRETIA	+15.4	-5.8	-3.7	-2.4
INSTITUUTIO	-2.2	+30.6	-7.9	-5.1
JOUKKO	-16.7	-5.9	+13.1	+53.6
YKSIÖ	-47.2	+34.6	+15.7	-0.4

Taulukko 24f. Kooste strategioista, jotka paikantavat merkitsevyyden sijaintia ja suuntaa¹¹⁰.

Strategia	Merkitsevyystaso	Luokka	ERI ¹¹¹	IRTI	KIRKKO	PUOLISO
Konservatiivinen	$X^2(df=12, a=0.05) > 21.026$	ELOTON	+	-	0	0
		KONKRETIA	0	0	0	0
		INSTITUUTIO	0	+	0	0
		JOUKKO	0	0	0	+
		YKSIÖ	-	+	0	0
Epäkonservatiivinen	$X^2(df=1, a=0.05) > 3.841$	ELOTON	+	-	-	-
		KONKRETIA	+	-	0	0
		INSTITUUTIO	0	+	-	-
		JOUKKO	-	-	+	+
		YKSIÖ	-	+	+	0
Standardoitu Pearsonin jäämä	e _{ij} /standardoitu Pearsonin jäämä > 2 ¹¹²	ELOTON	+	-	-	-
		KONKRETIA	+	-	-	0
		INSTITUUTIO	-	+	-	-
		JOUKKO	-	-	+	+
		YKSIÖ	-	+	+	0

Taulukossa 24f miinus- ja plusmerkit signaloivat merkitseviä havaintoja ja poikkeaman suuntaa. Plusmerkki tarkoittaa, että esiintymää on odotettua enemmän ja miinusmerkki että niitä on odotettua vähemmän. Nolla tarkoittaa, ettei havainto ole merkitsevä. Kun tarkastellaan konservatiivista osataulukkoa (jossa merkitsevät havainnot ovat $X^2 > 21.026$), voidaan tuloksia sanallistaa toteamalla, että luokka ELOTON on positiivisesti kytkeytynyt skeeman EROTA.ERI ($X^2=+98.5$) ja negatiivisesti skeeman EROTA.IRTI ($X^2=-37.1$). Luokka INSTITUUTIO assosioituu skeemaan EROTA.IRTI ($X^2=+30.6$), kuten jo ylempänä otaksuin. Epäkonservatiivisessa tarkastelussa (jossa merkitsevät havainnot ovat $X^2 > 3.841$) INSTITUUTIO on lisäksi negatiivisesti assosioitunut skeemoihin EROTA.KIRKKO ja EROTA.PUOLISO. Standardoitujen Pearsonin jäämien valossa (solun arvo > 2) negatiivinen kytkös ulottuu myös skeemaan EROTA.ERI. Luokka YKSIÖ on negatiivisesti kytkeytynyt skeemaan EROTA.ERI ja positiivisessa skeemaan EROTA.IRTI sekä epäkonservatiivisessa tarkastelussa lisäksi skeemaan EROTA.KIRKKO. Konservatiivisessa tarkastelussa luokka JOUKKO on positiivisesti kytköksissä skeemaan EROTA.PUOLISO; muut rivin havainnot eivät ole merkitseviä.

Satoja *post hoc*-taulukoita tutkineena olen havainnut tendenssin, johon Arppekin

¹¹⁰ Jälkikäteen tehtyjen *post hoc*-analyysien ongelma on, että ne nostavat dramaattisesti todennäköisyyttä sille, että hypoteesi (H_1) tulee hyväksytyksi sattumalta (Gries 2003: 81). Tyypin I virheen välttämiseksi jotkut tilastotieteilijät suosittelevat kriittistä merkitsevyystasoa modifioivan korjaimen, kuten *Bonferronin*-korjaimen, jota kielitieteelliseen aineistoon soveltaa Gries (mts. 81–82). Arppen tapaan en kuitenkaan käytä korjainta (ks. Arppen kritiikki ja perustelut 2008: 84–85). Bonferron-kritiikko Moran (2003: 405) kirjoittaa: “probability of finding some spurious results is quite high, but the chance of all the results being spurious is extremely improbable”. Koska työni X^2 -testitulokset lähes kautta linjan ovat selvästi merkitseviä, voidaan uskoa, että ne todella suurimmalta osin ovat merkitseviä. Suodatan tilastolliset tulokset joka tapauksessa päättelyni kautta.

¹¹¹ Tilan vuoksi notaatioita on jatkossa välillä lyhennetty muotoihin ERI, IRTI, HAV1., ABSTRK. ja niin edelleen.

¹¹² En esitä tutkimuksessani matemaattisia kaavoja. Kaavoja eksplikoivat Arppe (2008) ja Agresti (2002).

(2008: 99, 153) yhtyy ja joka näkyy myös taulukossa 24f¹³. Herkimmin pienimmätkin poikkeavuudet poimii standardoidut Pearsonin jäämät (61.6 %). Konservatiivinen tarkastelu antaa vähiten merkitseviä solukohtaisia havaintoja (35.8 %), kun taas epäkonservatiivinen tarkastelu (50.4 %) sijoittuu yleensä edellisten välille¹⁴. Kutsun konservatiivista tarkastelu herkkyytasoksi 3 (karkein), epäkonservatiivista herkkyytasoksi 2 (keskiverto) ja standardoituja Pearsonin jäämiä herkkyytaso 1 (herkin). Koska skeemojen väliset erot ovat toisiinsa mikroskooppisia, suosin herkkyytaso 1 (jota myös Arppe suosii). Se on strategioista kaikista kokeellisin ja tuottaa myös tuloksia, jotka muut tilastolliset lähestymistavat eivät aina pitäisi merkitsevinä tilastollisina poikkeamina.

Varsinaista assosiatiivisuuden voimakkuutta mitataan siihen kehitetyillä mittareilla (*summary measures of association*) (ks. Arppe 2008: 85–94, ks. myös Gries 2003). Kategorisiin muuttujiin voidaan soveltaa Arppen mukaan neljää mittaria, jotka ovat Cramérin V , Goodman-Kruskal Lambda (λ), Goodman-Kruskalin Tau (τ), Theilin uncertainty coefficient (U) (ks. Goodman ja Kruskal 1954; Theil 1970). Koska mittarien teoreettiset lähtökohdat eroavat, Arppe (mts. 92) painottaa, että ne myös mittaavat fundamentaalisesti eri asioita, eikä niitä siksi voi laittaa paremmuusjärjestykseen (vrt. kuitenkin Arppe mts. 429–498). Yhteistä mittareille on, että ne tutkivat taulukkoa kokonaisuutena ja antavat syötteeksi arvon väliltä 0–1. Mitä isompi arvo on, sitä voimakkaampaa on assosiaatio. Cramérin V on symmetrinen, loput ovat asymmetrisiä, mikä tarkoittaa, että mittari laskee kaksi arvoa, esimerkiksi arvot $\lambda_{\text{muuttuja|skeema}}$ ja $\lambda_{\text{skeema|muuttuja}}$. Jälkimmäinen ($\lambda_{\text{skeema|muuttuja}}$) tarkastelee assosiaation vahvuutta siten, että se olettaa skeemat muuttujien pohjalta ennustettavaksi riippuvuustekijäksi (*lexemes as predictable dependents*, Arppe mts. 96, 103). Edellisessä ($\lambda_{\text{muuttuja|skeema}}$) tilanne on päinvastainen: mitta-arvo indikoi, kuinka paljon skeeman tietäminen lisää varmuutta ennustaa muuttuja oikein (Arppe mts. 85–93, 155). Yksittäisen mittarin sijaan Arppe (2008: 90–94) suosittelee niiden käyttöä rinnan. Hyödynnän tutkimuksessani kolmea mittaria, jotka ovat Cramérin V (symmetrinen), λ :ta ja U :ta (epäsymmetrisiä).

Mittaustulosten sanalliseen tulkintaan Corbett ja Le Roy (2002: 189) ehdottavat seuraavia nyrkkisääntöjä: 0.0–0.10 = 'hyvin heikko', 0.10–0.19 = 'heikko', 0.20–0.29 = 'kohdalainen' ja 0.3–1 = 'vahva yhteys'. Ihmistieteissä (erityisesti) Cramérin V :n arvoja sanallis-

¹³ Työssäni käytettyjen taulukkojen ja mittarien tulokset on saatu R :ssä kirjoittamillani skripteillä, jotka hyödyntävät muun muassa *Polytomous*-kirjaston (Arppe 2015) komentoja *chisq.posthoc* ja *associations*.

¹⁴ Nämä prosenttiosuudet, jotka kuvastavat kuinka usein strategia löytää merkitseviä havaintoja, ovat Arppen työstä (2008), eivätkä sinällään liity tutkimukseeni. Ne havainnollistavat strategioiden eroja.

tetaan usein siten, että 0.3 = 'kohtalainen', 0.5 = 'melko voimakas' ja 0.8 = 'erittäin voimakas yhteys' (Heikkilä 2008: 213)¹¹⁵. Howellin (1999: 186) mukaan usein jo hyvin pienet arvot signaloivat huomionarvoista assosiaatiota (etenkin kielitieteessä)¹¹⁶.

Tulkinta on, että taulukon 24 assosiaation vahvuus ($V = 0.571$) on melko voimakasta (vrt. Arppe [2008: 107], jonka mukaan $V=0.397$ on *significant without a doubt*). Voidaan todeta, että muuttuja Y_SEM selittää noin reilun kolmanneksen ($\lambda_{\text{skeema|muuttuja}}=0.431$, $U_{\text{skeema|muuttuja}}=0.444$) skeemojen käyttäytymisestä (vrt. Arppe mts. 101–103, 155). Skeeman pohjalta taas voidaan selittää (tai "ennustaa") puolet ($\lambda_{\text{muuttuja|skeema}}=0.496$) tai ainakin reilu kolmannes ($U_{\text{muuttuja|skeema}}=0.389$) muuttujan Y_SEM käyttäytymisestä. Tämä tarkoittaa, että tietämällä, mistä skeemasta on kyse, voidaan ennustaa muuttujaluokka hieman paremmin kuin päinvastoin. Kytkös on siis vahvempaa skeemoista luokkiin kuin muuttujan luokista skeemoihin päin.

Lisätietoa *erota*-verbin agentin roolista merkitysvaihtelussa voidaan saada tutkimalla muuttujaa Y_MUOTO. Tällä kertaa X^2 -testin rajoitukset täyttyvät ilman luokkajärjestelyjä. Taulukko 25 esittää muuttujan jakauman ja kuva 25b havainnollistaa prosenttiosuuksia.

Taulukko 25. Muuttujan Y_MUOTO jakauma (n=558).

	ERI	IRTI	KIRKKO	PUOLISO	Σ_m
NP	138	93	39	37	307
NOLLA	8	24	32	17	81
INKORPOROITU	5	16	25	7	53
PERS	2	18	11	5	36
NA	1	11	6	6	24
PRON	52	1	3	1	57
Σ_s	206	163	116	73	558

Taulukko 25b. Muuttujan Y_MUOTO prosentuaalinen jakauma skeemojen välillä (vaalealla taustalla) ja skeemojen sisällä (tummalli taustalla).

	ERI		IRTI		KIRKKO		PUOLISO		Σ_m
NP	45.0	67.0	30.3	57.1	12.7	33.6	12.1	50.7	55.0 %
NOLLA	9.9	3.9	29.6	14.7	39.5	27.6	21.0	23.3	14.5 %
INKORPOROITU	9.4	2.4	30.2	9.8	47.2	21.6	13.2	9.6	9.5 %
PERS	5.6	1.0	50.0	11.0	30.6	9.5	13.9	6.8	6.5 %
NA	4.2	0.5	45.8	6.7	25.0	5.2	25.0	8.2	4.3 %
PRON	91.2	25.2	1.8	0.6	5.3	2.6	1.8	1.4	10.2 %
Σ_s	36.9 %		29.2 %		20.8 %		13.1 %		100 %

¹¹⁵ Arppe (2008: 94) muistuttaa, että sanalliset tulkintaohjeet ovat enemmän tai vähemmän summittaisia.

¹¹⁶ Cramérin V :tä hyödynnän, koska se on yleisesti käytössä kielitieteessä. Sen etuja on riippumattomuus otoksen koosta, mikä mahdollistaa erikokoisten näytteiden assosiaation vertailun (Gries & Stoll 2009: 1081). Lambdan (λ) valitsin sattumalta. Arppe (2008) suosii U :ta ja τ -ta – ja etenkin ensimmäistä. Lambdan λ :ta hän käyttää kai vertailukelpoisuuden takia, koska Gries (2003) käyttää sitä. Arppe (mts. 154) perustelee valintojaan näin: "Goodman-Kruskal τ as well as Theil's Uncertainty Coefficient U can both be considered to have better properties in comparison to the Goodman-Kruskal λ , as the formal two [τ ja U] take into account the entire distribution of possible outcome classes not only the mode as is the case the with the latter measure [λ]."

Taulukoita 25 ja 25b silmäilemällä havaitaan, että yleisin luokista on NP (55.1 %) ja harvinaisin NA (4.3 %). Sisäisesti NP:llä on isoin edustus skeemassa EROTA.ERI (66.7 %) ja pienin skeemassa EROTA.KIRKKO (33.6 %). Luokka PRON hakeutuu skeemojen välisessä vertailussa (91.2 %) skeemaan EROTA.ERI (n=52) ja karttaa skeemoja EROTA.IRT (n=1), EROTA.KIRKKO (n=3), EROTA.PUOLISO (n=1), mikä tarkoittaa, että skeemassa ERO.ERI esiintyy anaforisen viittaussuhteen omaavia *tämä*-, *se*- ja *joka*-edussanaisia subjektilausekkeita. Luokan INKORPOROITU painotus on suurinta skeeman EROTA.KIRKKO sisällä (21.6 %). Luokat PERS ja NA eivät ole järkeviä yleisiä missään skeemassa, joskin erityisesti skeemaa EROTA.ERI karttaa sitä (0.5 %) sekä lisäksi luokkaa NOLLA (3.9 %). Skeemojen välillä luokan NOLLA esiintymät keskittyvät skeemoihin EROTA.IRT (29.6 %), EROTA.KIRKKO (39.5 %) ja EROTA.PUOLISO (21.0 %), joista suurin sisäinen NOLLA-osuus on skeemalla EROTA.KIRKKO (27.6 %). Havainnot vahvistavat sitä jo todettua seikkaa, että skeemojen EROTA.IRT, EROTA.KIRKKO ja EROTA.PUOLISO agenteiksi hakeutuu elollisia olioita, kun taas skeeman EROTA.ERI agentti-tarkoituksena esiintyy elottomia tarkoituksia.

Solukohtaista merkitsevyyttä voidaan tarkastella Arppen tapaan plus- ja miinusmerkit sisältävällä taulukolla. Työssäni käytän kuitenkin niin sanottua *erotustaulukkoa*, jossa ilmenee kolme asiaa: havaintojen merkitsevyys, poikkeavuuksien suunta sekä teoreettisesti laskettujen odotettujen arvojen ja korpuksessa havaittujen frekvenssien välinen absoluuttinen erotus, joka antaa karkean käsityksen siitä, onko skeema kytköksissä luokkaan vai ei¹¹⁷. Absoluuttiset erotukset eivät kuitenkaan ole keskenään suoraan vertailukelpoisia, eikä niistä lasketa solukohtaista merkitsevyyttä (!) vaan ne lasketaan taulukoista (25x, ks. liite 1), jotka sisältää solukohtaiset X^2 -kertymät (herkkyystaso 3) ja standardoidut Pearsonin jäämät (herkkyystaso 1). Taulukko 25c esittää muuttujan Y_MUOTO erotustaulukon, jossa herkkyystaso 1 on vaalealla ja herkkyystaso 3 tummalla taustalla. Lihavoidut solut ovat merkitseviä.

Taulukko 25c. Muuttujan Y_MUOTO havaittujen ja odotettujen arvojen erotus ja suunta. Herkkyystasolla 1 (vaalea tausta) ja herkkyystasolla 3 (tumma tausta).

	ERI	IRT	KIRKKO	PUOLISO
NP	+ / 24.7	+ / 3.3	- / 24.8	- / 3.2
NOLLA	- / 21.9	+ / 0.3	+ / 15.2	+ / 6.4
INKORPOROITU	- / 14.6	+ / 0.5	+ / 14.0	+ / 0.1
PERS	- / 11.3	+ / 7.5	+ / 3.5	+ / 0.3
NA	- / 7.9	+ / 4.0	+ / 1.0	+ / 2.9
PRON	+ / 31.0 ¹¹⁸	- / 15.7	- / 8.8	- / 6.5

¹¹⁷ Esimerkiksi solun (EROTA.IRT:NP) arvo on laskettu kirjoittamani skriptin *erotus.R* kaavalla ”havaittu - odotettu arvo” eli $138-112.3 = (+) 24.7$. Plusmerkki tarkoittaa, että havaittuja on enemmän kuin odotettuja arvoja.

¹¹⁸ Herkkyystason 3 havainnot ovat merkitseviä tietysti myös herkkyystasolla 1.

Taulukon 25c tulokset tukevat jo todettuja päätelmiä, jotka ovat: i) NP on positiivisessa kytköksessä skeemaan EROTA.ERI ja negatiivisessa skeemaan EROTA.KIRKKO ja ii) NOLLA ja INKORPOROITU esiintyy odotettua enemmän skeemassa EROTA.KIRKKO (NOLLA = +15.2 esiintymää, INKORPOROITU = +14.0 esiintymää) ja odotettua vähemmän skeemassa EROTA.ERI (NOLLA = -21.9 ja INKORPOROITU = -14.6 esiintymää). Ainoa herkkyytasolla 3 merkitsevä havainto on yhteys luokan PRON ($X^2=+45.5$, ks. liite 1) ja skeeman EROTA.ERI (+31.0 esiintymää) välillä. Skeemojen EROTA.PUOLISO ja EROTA.IRTI havainnot eivät ole kiinnostavia. Yksi havainto on, että skeema EROTA.KIRKKO ehkä suosii nollapersoonaa ja inkorporoituja agentteja eli rakenteita, joihin puhuja voi sijoittaa itsensä (kertoakseen erokokemuksistaan).

Tulokseksi X^2 -testissä saadaan: $X^2=178.652$, $df=15$, $p < 2.2e^{-16}$. Tulkinta on, että taulukon solujen X^2 -jakaumien summa 178.652 on selvästi suurempi kuin khiin neliön kriittinen arvo, X^2 ($df=15$ ja $\alpha=0.05$) = 24.995. Myös p ($p = 2.2e^{-16}$) on selvästi kriittistä tasoa ($\alpha=0.05$) pienempi. Päätely on, että muuttujan Y_MUOTO ja skeemojen välillä on kytkösisyyttä, jonka voimakkuus on kohtalaista ($V=0.327$). Skeemat ennustavat vähän heikommin luokien ($U_{\text{muuttuja|skeema}}=0.126$, $\lambda_{\text{muuttuja|skeema}}=0^{119}$) kuin luokat skeemojen vaihtelua ($\lambda_{\text{skeema|muuttuja}}=0.198$, $U_{\text{skeema|muuttuja}}=0.132$), tosin yhteys luokista skeemoihin on heikohkoa.

5.2.2 Muuttujat Z_SEM ja Z_MUOTO

Tarkastelen seuraavaksi muuttujan Z_SEM suhdetta skeemoihin. Luokan NA (N=187) poistamisen jälkeen jäljelle jäi 371 esiintymää, joiden jakauma on esillä taulukossa 26. Taulukko 26b esittää prosentuaalisen jakauman sekä skeemojen kesken että skeemojen sisällä.

Taulukko 26. Muuttujan Z_SEM jakauma suhteessa skeemoihin.

	ERI	IRTI	KIRKKO	PUOLISO	\sum_m
ELOTON	113	2	0	1	116
KONKRETIA	21	0	0	0	21
INSTITUUTIO	20	91	95	0	206
JOUKKO	13	0	0	2	15
YKSIÖ	6	0	0	7	13
\sum_s	173	94	95	10	371

Taulukko 26b. Muuttujan Z_SEM (n=371) prosentuaalinen skeemojen välillä (vaalealla taustalla) ja sisällä (tumalla taustalla).

	ERI		IRTI		KIRKKO		PUOLISO		\sum_m
ELOTON	97.4	65.3	1.7	2.2	0.0	0.0	0.9	10.0	31.3
KONKRETIA	100.0	12.1	0.0	0.0	0.0	0.0	0.0	0.0	5.7
INSTITUUTIO	9.7	11.6	44.2	97.8	46.1	100.0	0.0	0.0	55.5
JOUKKO	86.7	7.5	0.0	0.0	0.0	0.0	13.3	20.0	4
YKSIÖ	46.2	3.5	0.0	0.0	0.0	0.0	53.8	70.0	3.5
\sum_s	46.6 %		25.1 %		25.6 %		2.7 %		100.0

¹¹⁹ Mittaustulos 0 tarkoittaa, että assosiaation voimakkuutta ei joko ole tai sitä ei pystytty laskemaan.

Silmiin pistävää on, että 187 esiintymää 558:sta eli 33.3 % on luokassa NA (joka ei ole mukana tarkastelussa), mistä seuraa, että 371 esiintymällä eli 66.6 % esiintyy ilmipantu rektioadverbiaalia. Taulukossa 26b havaitaan samanlaista systemaattisuutta tauluissa 25. Skeemassa ERO.ERI esiintyy 97.4 % luokan ELOTON ja 100.0 % luokan KONKRETIA esiintymistä. Yleisin luokka on INSTITUUTIO (n=206, 55.5 %), joka painottuu skeemoihin EROTA.IRTI (n=91) ja EROTA.KIRKKO (n=95)¹²⁰. Skeemassa EROTA.PUOLISO tapaa ilmipantua rektioadverbiaalia vain 10 kertaa. Luokat JOUKKO (n=15) ja YKSIÖ (n=13) ovat harvinaisia.

Taulukon 26 X^2 -testaus tuottaa ongelmia. Minimiehtojen kohta ”yksikään solu ei saa olla alle 1” ei täyty, vaikka luokat KONKRETIA ja ELOTON sekä JOUKKO ja YKSIÖ on yhdistetty, kuten ilmenee taulukosta 26c, joka esittää muuttujan Z_SEM odotetut frekvenssit. X^2 -testin tuloksiin tulee siksi suhtautua varauksella.

Taulukko 26c. Muuttujan Z_SEM odotetut frekvenssit.

	ERI	IRTI	KIRKKO	PUOLISO
ELOTON	63.8	34.3	35.0	3.7
INSTITUUTIO	96.0	51.7	52.7	5.5
ELOLLINEN	13.0	7.0	7.2	0.75

Koska X^2 -jaukaumien summa ($X^2=381.0528$, $df=6$, $p < 2.2e^{-16}$) on merkittävästi suurempi kuin khiin neliön kriittinen arvo, X^2 ($df=6$ ja $\alpha=0.05$) = 12.59, ja koska p ($p=2.2e^{-16}$) on selvästi kriittistä tasoa ($\alpha=0.05$) pienempi, voidaan päätellä, että muuttujien Z_SEM ja skeemojen välillä on hyvin todennäköisesti kytkösisyyttä, joka ei johdu sattumasta. Assosiaation voimakkuus on lähes erittäin voimakasta ($V=0.717$). Tietämällä, mistä skeemasta on kyse, pystytään lähes neljässä tapauksessa viidestä ($\lambda_{\text{muuttuja|skeema}}=0.745$, $U_{\text{muuttuja|skeema}}=0.599$) ennustamaan muuttujan luokka. Luokkien kytkös skeemoihin on sekin lähes melko voimakasta ($\lambda_{\text{skeema|muuttuja}}=0.379$; $U_{\text{skeema|muuttuja}}=0.464$).

Vaikka jo nyt on selvää, missä erot sijaitsevat, voidaan silti luoda silmäys taulukkoon 26d, joka esittää liitteen 1 taulukon 26x pohjalta määrittyvät solukohtaiset merkitsevyydet herkkyystasolla 1 (vaalea tausta) ja herkkyystasolla 3 (tumma tausta).

¹²⁰ Skeemassa EROTA.KIRKKO rektioadverbiaalin spesifioivat lekseemit: *joka* (n=1), *ahne kristillinen laitos* (n=1), *ne* (n=1), *seurakunta* (n=1), *uskonnollinen yhteisö* (n=1) ja *(evankelis-luterilainen) kirkko* (n=91). Skeeman EROTA.IRTI yleisimmät spesifikaatiot ovat: *euro* (n=14), *tehtävä* (n=6), *virka* (n=4), *EU* (n=4). Skeemat erottuivat toisistaan selvästi, jos INSTITUTION alakatégoriana olisi luokka USKONNOLLINEN INSTITUUTIO.

Taulukko 26d. Muuttujan Z_SEM havaittujen ja odotettujen arvojen erotus, suunta ja merkitsevyys kriteeritasolla 1 (vaalea tausta) ja herkkyystasolla 3 (tumma tausta). Merkitsevät solut on lihavoitu.

	ERI	IRTI	KIRKKO	PUOLISO
ELOTON	+ / 70.0	- / 32.4	- / 35.1	- / 2.7
INSTITUUTIO	- / 76.3	+ / 39.4	+ / 42.3	- / 5.5
ELOLLINEN	+ / 5.9	- / 7.0	- / 7.2	+ / 8.2

Taulukosta 26d erottuu kolme ryhmää. Skeema EROTA.ERI assosioituu positiivisesti luokkiin ELOTON (+70.0) ja JOUKKO (+5.9) ja negatiivisesti luokkaan INSTITUUTIO (-76.3). Skeemat EROTA.IRTI ja EROTA.KIRKKO käyttäytyvät täysin päinvastaisesti. Luokan ELOLLINEN ainut herkkyystason 3 löytö osuu skeemaan EROTA.PUOLISO. Löytöön on kuitenkin suhtauduttavat varauksella, koska nimenomaan tässä solussa odotettuja arvoja on alle 1. Koska skeeman ilmipannuista rektioadverbiaaleista (n=10) luokassa ELOLLINEN esiintyy 9 ja koska solulla on isoin X^2 -kertymä ($X^2=+90.8$, ks. liite 1), olen kuitenkin valmis uskomaan, että skeeman EROTA.PUOLISO rektioadverbiaaleiksi hakeutuu elollistarkoitteisia olentoja. Uskomuksen varmistamiseksi tarvittaisiin tutkimuskorpustani isompi populaatio.

Tuloksia voidaan syventää tarkastelemalla muuttujaa Z_MUOTO, jonka luokat järjestelin X^2 -testin miniehtojen vuoksi ryhmiin NA (\leftarrow NOLLA, n=4), NP (\leftarrow PERS, n=2) ja PRON (\leftarrow RAKENNE, n=11 ja PRON, n=23). Taulukko 27 esittää muuttujan jakauman yhdistämisen jälkeen. Taulukko 27b esittää prosenttijakaumat.

Taulukko 27. Muuttujan Z_MUOTO jakauma (n=558). Suurimmat arvot on lihavoituna.

	ERI	IRTI	KIRKKO	PUOLISO	\sum_m
NP	155	82	93	8	338
NA	33	70	21	63	187
PRON	18	11	2	2	33
\sum_s	206	163	116	73	558

Taulukko 27b. Muuttujan Z_MUOTO prosentuaalinen jakauma skeemojen välillä (vaalealla pohjalla) ja skeemojen sisällä (tummalla taustalla).

	ERI		IRTI		KIRKKO		PUOLISO		\sum_m
NP	45.9	75.2	24.3	50.3	27.5	80.2	2.4	11.0	60.6 %
NA	17.6	16.0	37.4	42.9	11.2	18.1	33.7	86.3	33.5 %
PRON	54.5	8.7	33.3	6.7	6.1	1.7	6.1	2.7	5.9 %
\sum_s	36.9 %		29.2 %		20.8 %		13.1 %		100.0 %

Taulukoita 27 ja 27b silmäilemällä havaitaan, että NP on yleisin luokka skeemoissa EROTA.KIRKKO (80.2 %), EROTA.ERI (75.2 %) ja EROTA.IRTI (50.3 %). Skeemassa EROTA.PUOLISO yleisin luokka kuitenkin on NA (86.3 %), johon myös skeemalla EROTA.IRTI (42.9 %) on taipumusta. Luokka PRON, jonka esiintymät skeemojen välillä painottuvat skeemoihin EROTA.ERI (54.5 %) ja EROTA.IRTI (33.3 %), ei vaikuta erityisen merkittävältä.

Koska taulukon 27 solujen X^2 -jakaumien summa ($X^2=146.890$, $df=6$, $p < 2.2e^{-16}$) on selvästi suurempi kuin khiin neliön kriittinen arvo, X^2 ($df=6$ ja $\alpha=0.05$) = 12.592, ja koska myös p ($p = 2.2e^{-16}$) alittaa ilmiselvästi kriittisen tason ($\alpha=0.05$), voidaan päätellä, että luokien ja skeemojen välillä on todennäköisesti riippuvuutta, jonka vahvuus on kohtalaista ($V=0.363$). Luokat selittävät noin kymmenesosan ($U_{\text{skeema|muuttuja}}=0.102$, $\lambda_{\text{skeema|muuttuja}}=0.104$) skeemojen vaihtelusta. Skeemojen kytkös luokkiin on hieman suurempaa ($U_{\text{muuttuja|skeema}}=0.160$, $\lambda_{\text{muuttuja|skeema}}=0.25$). Taulukko 27c tutkii poikkeavuuksien suuntaa esittämällä havaitun ja teoreettisesti odotetun jakauman erotuksen, poikkeavuuden suunnan ja soveltuisen merkitsevyyden herkkyystasoilla 1 ja 3 (arvot määrittyvät taulusta 27x, ks. liite 1)

Taulukko 27c. Muuttujan Z_MUOTO havaittujen ja odotettujen arvojen erotus, suunta ja merkitsevyys herkkyystasolla 1 (vaalea tausta) ja 3 (tumma tausta).

	ERI	IRTI	KIRKKO	PUOLISO
NP	+30.0	-16.7	+22.7	-36.2
NA	-36.0	+15.4	-17.8	+38.5
PRON	+5.8	+1.4	-4.9	-2.3

Taulukosta 27c erottuu kaksi ryhmää. Ensimmäinen hakeutuu lauserakenteisiin, joissa on ilmiäntuna rektioadverbiaali. Tähän kuuluvat EROTA.ERI (NP = +30) ja EROTA.KIRKKO (NP = +22.7). Toinen ryhmä päinvastoin hakeutuu lauserakenteisiin, jotka suosivat rektioadverbiaalin poisjättämistä. Tähän kuuluvat EROTA.IRTI (NA = +15.4) ja etenkin EROTA.PUOLISO (NA = +38.5), joka on herkkyystason 3 löydöistä painavin ($X^2 = +60.7$, ks. liite 1). Kiinnostavaa on verrata tulosta muuttujan Z_SEM tuloksiin, jossa semanttisesti läheisten skeemojen EROTA.IRTI ja EROTA.KIRKKO esiintymät jakautuivat lähes identtisesti luokkaan INSTITUUTIO (EROTA.IRTI, $n=91$; EROTA.KIRKKO, $n=95$) – Y_SEM:NA-luokan poistamisen takia tuloksessa ei ilmennyt, että skeemalla EROTA.IRTI on isompi otos. Tulos hämärtää sen tosiasian, että skeemalla EROTA.KIRKKO esiintymistä elatiivisijainen adverbiaali (*kirkko*) on ilmiäntuna 81.9 % esiintymistä, kun skeemalla EROTA.IRTI ilmiäntuja rektioadverbiaaleja on 57 %. Vahvin suhde rektioadverbiaalin poisjättämiseen on kuitenkin skeemalla EROTA.PUOLISO (NP = -36.2, NA=+38.5). Tulkinta on, että parisuhteen päättämisen merkitys on siinä määrin kontekstin (leksikaalisten ja temaattisten) vihjeiden perusteella pääteltävissä oleva käsitteistys, ettei ”merkityksen” alleviivaamiseen tarvita agenttia tai lähettä. Osittain päinvastainen on skeeman EROTA.KIRKKO merkityksenmuodostus: luenta syntyy nimenomaan lähteen ilmiäntemisestä lauserakenteeseen.

5.2.3 Muuttujat RAKENNE ja RESIPROOKKI

Muuttuja RAKENNE mittaa, esiintyvätkö Y- ja Z-argumentit ilmauksissa. Se mittaa siis osit-

tain samoja asioita kuin jo käsitellyt muuttujat, mutta kuitenkin sillä erolla, että se ottaa huomioon Y- ja Z-argumenttien yhteisvaikutuksen. *erota*-verbin yhteydessä muuttujaa saa seitsemän ala-arvoa, jotka ovat VAIN_Y, Y_JAOLLINEN, VAIN_Z, REKTIO, REKTIO2, KONNEKTORI ja NA. X^2 -testausta varten yhdistin mikroluokan REKTIO2 (n=1) luokkaan REKTIO (n=288)¹²¹. Taulukko 28 esittää muuttujan jakauman. Taulukko 28b esittää skeemojen sisäiset (tumma taustalla) ja skeemojen väliset (vaalealla taustalla) prosenttijakaumat.

Taulukko 28. Muuttujan RAKENNE jakauma. Isoimmat havainnot on lihavoitu (n=558).

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO	Σ_m
REKTIO	146	70	65	5	286
VAIN_Y	7	55	13	15	90
VAIN_Z	6	23	31	5	65
Y_JAOLLINEN	25	5	0	31	61
NA	1	9	7	17	34
KONNEKTORI	21	1	0	0	22
Σ_s	206	163	116	73	558

Taulukko 28b. Muuttujan RAKENNE prosentuaalinen jakauma skeemojen sisällä (tumma tausta) ja skeemojen välillä (vaalea tausta).

	ERI		KIRKKO		IRTI		PUOLISO		Σ_m
REKTIO	51.0	70.9	22.7	56.0	24.5	42.9	1.7	6.8	51.3
VAIN_Y	7.8	3.4	14.4	11.2	61.1	33.7	16.7	20.5	16.1
VAIN_Z	9.2	2.9	47.7	26.7	35.4	14.1	7.7	6.8	11.6
Y_JAOLLINEN	41.0	12.1	0.0	0.0	8.2	3.1	50.8	42.5	10.9
NA	2.9	0.5	20.6	6.0	26.5	5.5	50.0	23.3	6.1
KONNEKTORI	95.5	10.2	0.0	0.0	4.5	0.6	0.0	0	3.9
Σ_s	36.9 %		20.8 %		29.2 %		13.1 %		100.0

Taulukkoja 28 ja 28b silmäilemällä havaitaan, että yleisimmällä luokalla REKTIO:lla (51.0 %) – joka vastaa argumenttirakennetta SVA_{ero} – on sisäinen valta-asema skeemoissa EROTA.ERI (70.9 %), EROTA.KIRKKO (56 %) ja EROTA.IRTI (42.9 %). Ainoastaan skeemassa EROTA.PUOLISO sen osuus on pieni (REKTIO = 6.9 %); siinä valtaa 43.1 %:lla pitää Y_JAOLLINEN (*pari erosi*), jonka esiintymät skeemojen välillä painottuvat skeemoihin EROTA.PUOLISO (51.7 %) ja EROTA.IRTI (40 %). Skeemojen EROTA.IRTI ja EROTA.KIRKKO välillä näkyy eroa luokkien VAIN_Y (EROTA.IRTI = 33.7 % , EROTA.KIRKKO = 11.2 %) ja VAIN_Z (EROTA.IRTI = 14.1 % , EROTA.KIRKKO = 26.7 %) skeeman sisäisissä painotuksissa, joiden esiintymät keskittyvät myös skeemoihin EROTA.KIRKKO (VAIN_Z = 47.7 %) ja EROTA.IRTI (VAIN_Y = 61.1 %).

Skeemojen välisessä tarkastelussa havaitaan, että luokan KONNEKTORI esiintymät painottuvat skeemaan EROTA.ERI (95.5 %); muissa skeemoissa sitä ei juuri esiinny. Skeeman EROTA.PUOLISO esiintymistä taas 86.3 % (VAIN_Y = 20.5 % + NA = 23.3 % + Y_JAOLLINEN

¹²¹ Yhdistelyn jälkeen alle 5 esiintymää on 3 solussa, mutta testin ehdot täyttyvät, sillä 87.5 % arvoista on > 5.

= 42.5 %) on argumenttirakenteissa, joissa ilmirakenteessa esiintyy pelkästään agentti (63.0 %) tai finiittiverbi(ketju) paljaana ilman agenttia ja lähdetä (23.3 %). Tulos vahvistaa muuttujan Z_MUOTO yhteydessä tehtyä havaintoa.

Koska taulukon 28 testisuure ($X^2 = 306.537$, $df=15$, $p < 2.2e^{-16}$) on selkeästi suurempi kuin khiin neliön kriittinen arvo X^2 , ($df=12$, $\alpha=0.05$) = 24.996, ja koska myös p ($X^2 = 306.537$, $df=15$, $p > 2.2e^{-16}$) alittaa riskitason ($\alpha=0.05$), voidaan todeta, että skeemojen ja muuttujan RAKENNE välillä on melko voimakasta kytkösisyyttä ($V=0.428$). Tällä kertaa luokat selittävät noin neljäsosan ($U_{\text{skeema|muuttuja}}=0.209$, $\lambda_{\text{skeema|muuttuja}}=0.270$) skeemojen vaihtelusta, kun taas skeemojen pohjalta voidaan ennustaa ($U_{\text{muuttuja|skeema}}=0.194$, $\lambda_{\text{muuttuja|skeema}}=0.096$) vain alle viidesosa luokkien vaihtelusta.

Lopuksi tutkin, missä erot sijaitsevat. Taulukko 28b esittää havaitun ja odotetun jakauman erotuksen, solukohtaisen merkitsevyyden ja suunnan (jotka määrittyvät liitteen 1 taulukon 28x pohjalta).

Taulukko 28b. Muuttujan RAKENNE havaittujen ja odotettujen arvojen erotus ja suunta. Merkitsevät solut herkkyystasolla 3 (tumma tausta) ja herkkyystasolla 1 (vaalea tausta). Merkitsevät solut on lihavoitu.

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO
REKTIO	+ / 40.4	- / 13.4	+ / 5.5	- / 32.4
KONNEKTORI	+ / 12.9	- / 5.4	- / 4.6	- / 2.9
VAIN_Z	- / 18.0	+ / 4.0	+ / 17.5	- / 3.5
Y_JAOLLINEN	+ / 2.5	- / 12.8	- / 12.7	+ / 23.0
NA	- / 11.6	- / 0.9	- / 0.1	+ / 12.6
VAIN_Y	- / 26.2	+ / 28.7	- / 5.7	+ / 3.2

Taulukosta 28b ilmenee neljä kytköstä herkkyystasolla 3 (tilastollisesti merkitsevin). Kolme niistä on skeemassa EROTA.PUOLISO, jolla on selvästi yksilöllinen profiili. Se hylkii luokkaa REKTIO (-32.4) ja suosii ainoana skeemana luokkia Y_JAOLLINEN (+23) ja NA (-12.6), jotka vastaavat argumenttirakenteita $S_{\text{jaollinen}V}$ (*Pari erosi*) ja V (*kannattaa erota*). Nimenomaan kytkös luokkaan Y_JAOLLINEN ($X^2=69.871$, ks. liite 1) on erityisen merkitsevä. Lähes täydellisenä vastaparina toimii skeema EROTA.ERI, joka suosii luokkia REKTIO (+40.4) ja KONNEKTORI (12.9) ja hylkii luokkia VAIN_Z, NA ja VAIN_Y – siis argumenttirakenteita joissa sekä Y- ja Z-argumentti ovat ilmipantuina! Neljäs herkkyystason 3 havainto skeeman EROTA.IRTI kytkös luokkaan VAIN_Y, joka vastaa argumenttirakennetta SV (*Ministeri erosi*). Skeema EROTA.IRTI liittyvät herkkyystason 1 löydöt signaloivat negatiivista kytköstä luokkiin REKTIO, KONNEKTORI ja Y_JAOLLINEN. Ainoastaan skeema EROTA.KIRKKO on positiivisessa assosiaatiossa luokkaan VAIN_Z (+17.5). Skeeman EROTA.IRTI kytkös luokkaan VAIN_Z ei poikkea odotuksesta (+3.2 esiintymää, $X^2=0.541$, ks. liite 1): ei siis voida väittää, että skeema EROTA.IRTI hylkisi luokkaa VAIN_Z.

Lopuksi tarkastelen muuttujaa RESIPROOKKI, jonka jakauma on esillä taulukossa 29. Poikkeuksellisesti esillä on kaikki *erota*-verbin skeemat. Taulukko 29b esittää muuttujan RESIPROOKKI ja RAKENNE ristiintaulukoinnin, joka paljastaa muuttujien välisen assosiaation.

Taulukko 29. Muuttujan RESIPROOKKI jakauma (n=563).

	ERI	IRTI2	ABSTRAKTI	KONKRETIA	PUOLISO	IRTI	KIRKKO	Σ_f
NA	177	7	2	2	73	155	116	532
RESIPROOKKI	29	1	1	0	0	0	0	31

Taulukko 29b. Muuttujan RESIPROOKKI ja NA ristiintaulukointi (n=563).

	Y_JAOLLINEN	KONNEKTORI	VAIN_Z	NA	VAIN_Y	REKTIO
NA	48	6	65	35	90	286
RESIPROOKKI	14	17	0	0	0	0

Taulukot puhuvat tällä kertaa puolestaan: muuttujaa RESIPROOKKI esiintyy lähes pelkästään skeemassa EROTA.ERI (n=29). Lisäksi sitä esiintyy kerran skeemoissa EROTA.IRTI2 (*kirkko ja valtio eroavat toisistaan*) ja EROTA.ABSTRAKTI (*kädellisten ja lihansyöjien linjat erosivat toisistaan*). RESIPROOKKI kytketty rakenteisiin Y_JAOLLINEN (*Kissat eroavat toisistaan*) ja KONNEKTORI (*Ankat ja oravat eroavat toisistaan*), mikä on odotuksenmukaista, koska *toisistaan*-lekseemi asettuu elatiivisijaisen täydennyksen paikalle. Tulkinta on, että RESIPROOKKI vaatii lauseeseen tasapuolisesti vastavuoroistuvat osallistujat¹²², joista kumpikaan ei ole toista ”ensisijaisempi”. Tällainen on nimenomaan skeeman EROTA.ERI hahmotama käsitteistys ominaisuuksien eroavaisuudesta.

5.2.4 Yhteenveto

erota-verbin yksimuuttujainen analyysi osoittaa, että verbin ydinargumentteja analysoimalla on mahdollista selvittää merkityksenvaihtelun mekaniikan pääpiirteitä. Tulokset osoittavat, että skeemojen välistä vaihtelua selittävät parhaiten muuttujat, jotka mittaavat ydinargumenttien semanttista ontologiaa. Tulos on siinä mielessä odotuksenmukainen, että tilastoyksiköiden annotointi skeemoihin on tehty semanttisin perustein intuition pohjalta. Olisi kummallista, jos ryhmien välillä ei löytyisi semanttisia eroja.

Sitä, mitkä muuttujien alaluokista ovat voimakkaimmin kytköksissä skeemoihin havainnollistaa taulukko 30. Tähän taulukkoon olen Arppen hengessä (2008: 157) seulottu 24:stä yksimuuttujaisessa tarkastelussa mukana olleesta kategoriasta 10 sillä perusteella,

¹²² Resiprookista adverbialia ei esiinny skeemoissa EROTA.IRTI~ KIRKKO, koska ne manifestoivat henkilötarkoitteisen olion SEPARAATIO:ta LÄHDEVÄYLÄ:ä pitkin. Tällöin fokuksena on olion tilan muutos suhteessa lähteeseen. Siksi lause *Sipilä ja pääministerin työ erosivat toisistaan* on omituinen, joskaan ei epäkielellinen. On sen sijaan helppo keksiä skeeman EROTA.PUOLISO toteutumiksi resiprookki-lauseita, kuten *Puolisot erosivat toisistaan*, joihin tosin helposti tulee spatiaalista sävyä (*Puolisot erosivat toisistaan pihalla*).

kuinka voimakkaasti kolmen eri assosiaatiomittarin nojalla i) yksittäiset luokat (ID merkki) ja skeemat ovat kytköksissä (Cramérin V), ii) skeemat ovat kytköksissä yksittäisiin luokkiin ($U_{\text{muuttuja|skeema}}$) ja iii) luokat ovat kytköksissä skeemoihin ($U_{\text{skeema|muuttuja}}$)¹²³.

Taulukko 30. Yksittäisten luokkien assosiaation voimakkuus skeemoihin tarkasteltuna kolmella mittarilla.

Cramérin V (≥ 0.400)	$U_{\text{muuttuja skeema}}$ (≥ 0.2)	$U_{\text{skeema muuttuja}}$ (≥ 0.06)
Z_SEM_INSTITUUTIO (0.886)	Z_SEM_INSTITUUTIO (0.719)	Z_SEM_INSTITUUTIO (0.430)
Z_SEM_ELOTON (0.786)	Z_SEM_ELOTON (0.569)	Z_SEM_ELOTON (0.326)
Y_SEM_ELOTON (0.747)	Y_SEM_ELOTON (0.560)	Y_SEM_ELOTON (0.270)
Y_SEM_YKSIÖ (0.575)	Z_SEM_ELOLLINEN (0.364)	Y_SEM_YKSIÖ (0.150)
Z_SEM_ELOLLINEN (0.558)	Y_SEM_YKSIÖ (0.303)	Z_MUOTO_NA (0.095)
Y_SEM_JOUKKO (0.510)	RESIPROOKKI_NA (0.232)	Y_SEM_JOUKKO (0.093)
Z_MUOTO_NA (0.498)	RESIPROOKKI_LÄSNÄ (0.232)	Z_MUOTO_NP (0.085)
Z_MUOTO_NP (0.463)	RAKENNE_Y_JAOLLINEN (0.230)	Z_SEM_ELOLLINEN (0.084)
RAKENNE_Y_JAOLLINEN (0.422)	Y_SEM_KONKRETIA (0.220)	RAKENNE_REKTIO (0.072)
RAKENNE_REKTIO (0.412)	Y_SEM_JOUKKO (0.219)	RAKENNE_Y_JAOLLINEN (0.060)

Kaikkien kolmen mittarien valossa voimakkainta assosiaatioita osoittavat luokat Z_SEM_INSTITUUTIO ($V=0.886$), Z_SEM_ELOTON ($V=0.786$), Y_SEM_ELOTON ($V=0.746$), Y_SEM_YKSIÖ ($V=0.575$) sekä Z_SEM_ELOLLINEN ($V=0.558$), joka on näistä heikoimmin kytköksissä skeemojen suuntaan ($U_{\text{skeema|muuttuja}}=0.084$). Assosiaatio on kautta linjan heikompaa muuttujaluokista (ID-merkeistä) skeemoihin päin kuin toiseen suuntaan, mikä tarkoittaa, että skeemojen pohjalta on helpompi ennustaa muuttujan luokka kuin päinvastoin.

Yksimuuttujaisen analyysin perusteella, skeemojen sisäisiin prosentiosuuksiin nojaten luokka Z_SEM_INSTITUUTIO – joka tarkoittaa, että ilmipannun lähteen spesifioi luokan INSTITUUTIO tarkoite – kytkeytyy skeemoihin EROTA.KIRKKO (100 % *Jeesus ei ole eronnut kirkosta*) JA EROTA.IRTI (97.8 %), kun taas luokka Y_SEM_YKSIÖ (agentti on yksilötarkoitteinen) kytkeytyy skeemaan EROTA.IRTI (*Sipilä erosi*). Luokat Y_SEM_ELOTON ja Z_SEM_ELOTON – jotka tarkoittavat, että sekä agentti että lähde ovat elottomia – poimivat skeeman EROTA.ERI (*Sotapuheet eroavat sotatoimista*) erilleen. Elollista lähettä tarkoittava luokka Z_SEM_ELOLLINEN (*En ole eronnut vaimostani*) sekä kollektiivitarkoitteista agenttia tarkoittava luokka RAKENNE_Y_JAOLLINEN ($V=0.422$) (*Lesboparit eroaa herkemmin*) kytkeytyvät erityisesti skeemaan EROTA.PUOLISO. Luokka RAKENNE_REKTIO (joka signaloi argumenttirakennetta SVA_{ero}) kytkeytyy eritoten skeemaan EROTA.ERI.

ELOTON : ELOLLINEN -jaottelun rinnalla jo rintaman YKSIÖ : JOUKKO : INSTITUUTIO sisällä on semanttisia eroja, jotka luokitusysteemini karkeus häivyttää. Niitä havainnollistaa

¹²³ Listaus poikkeaa edellisen luvun ristiintaulukoinneista siinä, että niissä on kerralla tutkittava kaikki muuttujan alakategoriat. Luokkia on mahdollista tarkastella myös erikseen suhteessa muihin muuttujan luokkiin, jolloin saadaan selville, kuinka paljon kukin luokka kytkeytyy skeemoihin (näin tekee Arppe 2008).

taulukko 30a, johon on poimittu skeemakohtaisesti yleisimmät Y- ja Z-argumenttien spesifikaatiot luokista YKSILÖ, JOUKKO ja INSTITUUTIO.

Taulukko 30a. Luokkien YKSILÖ, JOUKKO ja INSTITUUTIO yleisimmät spesifikaatiot.

	EROTA.ERI	EROTA.IRTI	EROTA.KIRKKO	EROTA.PUOLISO
Y_SEM_ YKSILÖ ~ JOUKKO	<i>ihminen</i> (n=4), populaatio kuten <i>suomalaiset</i> (n=3), <i>hän ~ he ~ ne</i> (n=3), <i>poliisi</i> (n=1), <i>älykäs henkilö</i> (n=1), <i>minä</i> (n=1); <i>karhut</i> (n=1), yht. n=25	nimetty henkilö esim. <i>Vanhanen</i> (n=25), ammattinimike esim. <i>ministeri</i> (n=16), (<i>pää- yms.</i>) <i>johtaja</i> (n=13), <i>hän ~ he ~ ne</i> (n=13), <i>minä</i> (n=7), <i>itse</i> (n=5), <i>sinä</i> (n=2), <i>me</i> (n=2), yht. n=96	<i>minä</i> (n=23), <i>lapsi ~ 15-17-vuotias ~ nuori</i> (n=6), <i>moni ~ usea ~ kaikki</i> (n=6), (4000 yms.) <i>ihmistä</i> (n=6) <i>itse</i> (n=5), <i>sinä</i> (n=3), <i>vanhemmat</i> (n=3), yht. n=78	<i>vanhemmat</i> (n=11), (<i>avio</i>) <i>pari</i> (<i>skunta</i>) (n=8), <i>nainen</i> (n=3), <i>hän ~ he</i> (n=6), <i>isä ~ äiti</i> (n=2), <i>minä</i> (n=2), <i>sinä</i> (n=1), <i>lesbot</i> (n=2), yht. n=51
Z_SEM_ YKSILÖ ~ JOUKKO	ryhmä esim. <i>länsisuomalaiset</i> (n=3), <i>eläin</i> (n=2) <i>rikollinen</i> (n=1), <i>sinä</i> (n=1)	-	-	<i>vaimo ~ emäntä</i> (n=4), <i>äiti</i> (n=1), <i>mies</i> (n=1)
Z_SEM_ INSTIT.	maa esim. <i>Suomi</i> (n=9), <i>YLE</i> (n=1), <i>SMP-liike</i> (n=1)	<i>tehtävä ~ työ yms.</i> (n=20), <i>euro</i> (n=17), <i>EU</i> (n=6), <i>valtio</i> (n=4), <i>kirkko</i> (n=1)	(<i>ev.lut yms.</i>) <i>kirkko</i> (n=90), <i>seurakunta yms.</i> (n=3)	-

Taulukosta ilmenee tendenssejä. Skeeman EROTA.PUOLISO argumentteina esiintyy nimenomaan naisiin ja perheenjäseniin viittaavia sanoja, kuten *nainen*, *äiti*, *lesbot* ja *vanhemmat*. Skeeman EROTA.ERI spesifikaatioina esiintyy paitsi pronomineja, kuten *minä*, joka poimii tarkoitteodellisuudesta definiittisen kohteen, usein myös kategorioiden ylätasoon nimityksiä, kuten *älykäs henkilö* (’joku kategoriasta älykäs henkilö’), *ihminen* (’joku kategoriasta ihminen’ tai ’ihminen lajina’) tai *suomalaiset* (’Suomen kansa’). Tarkoite voi myös olla eläin. Skeemassa EROTA.IRTI esiintyy paljon yleisnimiä, kuten *Vanhanen* tai muunlaisia definiittisiä kohteita ([tietyn] *yhtiön pääjohtaja* tai *me* [’Suomi valtiona’]). Lisäksi skeeman spesifikaatioita toimii definiittisiä pronomineja *minä* ja *hän*. Skeema EROTA.KIRKKO suosii ihmisryhmiin viittaavia sanoja, jotka voivat rajata joukon alan ”epämääräisesti” (*moni* ja *usea*) tai ”täydellisesti” (*kaikki* tai *4000 ihmistä*) (vrt. Leino 1993: 230). Lisäksi sen spesifikaatioita toimii ikäkauteen viittaavia sanoja (*lapsi* tai *15–17-vuotias*), joskin valta-asema on puhujan itseensä viittaavalla tyypillä *minä*. Luokan Z_SEM_INSTITUUTIO havainnot jakautuvat siten, että skeema EROTA.KIRKKO suosii lähes yksinomaan lekseemiä *kirkko*, skeema EROTA.IRTI useanlaisia sanoja, kuten *tehtävä*, *työ*, *euro*, *EU* ja *valtio*. Skeema EROTA.ERI voi spesifoida kaikenlaisia tarkoitteita, mutta aineistossani etenkin maan nimityksiä, kuten *Suomi*.

Kiinnostavaa on lisäksi pohtia sitä, miten pelkkien skeemojen välisten prosentiosuuk-sien pohjalta, voidaan Griesin (2006: 83–85) hengessä Gildean ja Jurafskyn (2001) tapaan tutkia, mitkä luokat tai luokkien yhdistelmät ”ennustavat” tiettyä skeemaa aidossa kontekstissa (tai kuten Gries [mts. 84] kysyy, ”how well can we predict the sense of [verbi] in a particular citation when we extract this citation’s ID tags?”). Tällöin on kyse tietokoneling-

vistiikassa yleisestä kysymyksestä, nimittäin automaattisesta merkityksen erottamisesta (*automatic sense identification* tai *word sense disambiguation*, WSD; ks. esim. Manning ja Schütze 2000: Kappale 7). Voidaan kuvitella tilanne, jossa alamerkitysten tunnistamisalgoritmiin on syötetty kyky tunnistaa *erota*-lekseemin seuralaisten semanttisia ja rakenteellisia piirteitä sekä lisäksi tieto siitä, millaisen prosenttijakauman nojalla ne esiintyvät kunkin *erota*-skeeman yhteydessä. Tällöin tarkastellaan skeeman tunnistamismahdollisuutta p (alamerkitys | ID merkki) vastakohtana tunnistamismahdollisuudelle p (ID merkki | alamerkitys), joita heijastavat assosiaatiomittareissa oppositio $U_{\text{muuttuja|skeema}} : U_{\text{skeema|muuttuja}}$. Taulukko 30b osoittaa, miten skeemojen välisten prosenttisuuksien kärki jakautuu korpuksesani *erota*-skeemoihin.

Taulu 30b. *erota*-skeemoja ennustavat luokat.

Skeema	"Ennustus"	Muuttujan luokka	Luokan kuvaus
EROTA.ERI	100 % 97.8 % 96.7 % 95.5 % 91.2 % 86.7 % 51.0 %	Y_SEM_ELOTON Z_SEM_ELOTON RESIPROOKKI_LÄSNÄ RAKENNE_KONNEKTORI Y_MUOTO_PRON Z_MUOTO_JOUKKO RAKENNE_REKTIO	agentin tarkoite eloton lähteen tarkoite eloton täydennyksenä esiintyy <i>toisistaan</i> -lekseemi argumenttirakennetyypinä "Y ja Z eroavat" agentin spesifioi pronomini (<i>tämä, se</i> yms.) lähde joukkotarkoitteen argumenttirakennetyypinä SVA _{ero}
EROTA.IRTI	71.1 % 61.6 % 55.2 % 50.0 % 45.8 % 37.4 %	Y_SEM_INSTITUUTIO RAKENNE_VAIN_Y Y_SEM_YKSILÖ Y_MUOTO_PERS Y_MUOTO_NA Z_MUOTO_NA	agentin tarkoite instituutio argumenttirakennetyypinä SV agentin tarkoite yksilö agenttina persoonapronomini agentti ei ilmipantuna lähde ei ilmipantuna
EROTA.KIRKKO	100 % ¹²⁴ 47.7 % 47.2 %	Y_SEM_ELOTON ~ INSTITUUTIO RAKENNE_VAIN_Z Y_MUOTO_INKORP.	agentti on eloton tai instituutio/organisaatio argumenttirakenne tyyppiä (nolla)VA _{ero} agentti inkorporoitu verbiin
EROTA.PUOLISO	100 % 53.8 % 50.8 % 50.0 %	Y_SEM_ELOTON ~ INSTITUUTIO Z_SEM_YKSILÖ RAKENNE_Y_JAOLLINEN RAKENNE_NA	agentti on eloton tai instituutio/organisaatio lähde yksilötarkoitteen argumenttirakenne tyyppiä S _{jaollinen} V argumenttirakenne tyyppiä V

Erityisen helppoa on paikantaa skeema EROTA.ERI. HS.fi-korpuksessa 100 %:n (n=136) erotelutarkkuus saavutetaan¹²⁵, jos algoritmi havaitsee, että ilmipannun agentin (Y-argumentin) tarkoite on ELOTON ~ KONKRETIA ja 97.8 %:n tarkkuus, jos ilmipantu lähde (Z-argumentti) kuuluu luokkaan ELOTON ~ KONKRETIA. Muuttujan RESIPROOKKI eli *toisistaan*-lekseemin läsnäolo antaa aineistossani 96.7 % (n=29) tarkkuuden, kun taas ilmipannun agentin (Y-argumentin) luokkaan PRON (*tämä, se* yms.) tunnistaminen antaa jakauman, jossa 91.2 % (n=52) lauseista on skeeman EROTA.ERI toteutumia. Kategorian RAKENNE_KONNEKTORI –

¹²⁴ Huutomerkki tarkoittaa prosenttijakaumaa, jonka nojalla kyseessä ei voi olla kyseinen skeema (0 %).

¹²⁵ Erottelutarkkuudella en tarkoita skeeman oikein ennustamisen todennäköisyyttä vaan skeemojen välistä prosenttijakaumaa, joka mainituilla ehdoilla kytkeytyy tutkimuskorpuksessani tiettyyn skeemaan. Ensimmäisessä esimerkissä erottelutarkkuuden notaatio "100 %:n (n=136)" tarkoittaa siten, että ehdot (patientti on eloton) pätevät 136:een skeeman EROTA.ERI korpusesimerkkiin. 100 %:n tarkkuus tarkoittaa, etteivät ehdot päde minkään muun skeeman esiintymiin. 80 %:n tarkkuus tarkoittaisi, että annetuilla ehdoilla 80 % korpusesimerkeistä sijoittuu skeemaan EROTA.ERI ja 20 % muihin skeemoihin.

joka tarkoittaa, että merkitysrakenteeseen kuuluvat jäsenet rinnastetaan *ja*-konnektorin tai vastaavan avulla – läsnä ollessa 95.5 % (n=21) korpusesimerkeistä kuuluu skeemaan EROTA.ERI. Asiaa voi katsoa toisinkin päin. Jos algoritmi havaitsee, että lause on passiivissa tai että *erota*-lekseemin ydinargumenttina ei esiintyy rektioadverbiaalia (Z-argumenttia) ja että agentin spesifioi YKSIÖ, voidaan 100 %:n tarkkuudella todeta, ettei kyse ole skeemasta EROTA.ERI, sillä kyseiset ehdot eivät koskaan esiinny HS.fi-aineistossa skeeman EROTA.ERI toteumissa. Toisaalta jos algoritmi tunnistaa, että argumenttirakenne on tyyppiä SVA_{ero} (*Y eroaa Z:sta*) ja että agentin spesifioi YKSIÖ, saadaan jakauma, jossa 88 %:ssa (n=72) lause-esimerkeistä on kyse joko skeemasta EROTA.KIRKKO tai EROTA.IRTI. Jos lisäksi havaitaan, että rektioadverbiaalin spesifioi uskonnollinen INSTITUUTIO, voidaan lähes 100 %:ssa tapauksissa todeta, että kyseessä on skeema EROTA.KIRKKO. Jos algoritmi sen sijaan tunnistaa, että ilmipantu agentti on joukkotarkoitteinen sekä lisäksi muuttujan Y_JAOLLINEN hengessä jaollinen ja että ydinargumenttina ei esiinny rektioadverbiaalia, saavutetaan jakauma, jossa 91.2 % (n=31) lukeutuu skeemaan EROTA.PUOLISO. En jatka spekulatiota, vaan esitän lopuksi yksimuuttujaisen analyysin tulosten pohjalta *erota*-skeemoille luonnehdinnat, joiden määrittely ensisijaisesti nousee määrittelysuunnasta *p* (ID-merkki eli LUOKKA | SKEEMA).

Skeeman EROTA.ERI korpushavainnoista 70.9 % on argumenttirakenteessa SVA_{ero}, jossa ilmipantuina ovat sekä agentti (Y-argumentti) että lähde (Z-argumentti) ja jossa kummankin näistä spesifioi ELOTON tarkoite, esimerkiksi *pohjoiskarjalan murre* (Y_SEM_ELOTON~KONKRETIA = 80.7 %; Z_SEM_ELOTON~KONKRETIA = 77.4 %). Aineistossani nämä kriteerit läpäisee 51 % (n=106) skeeman EROTA.ERI havainnoista, kuten esimerkiksi *Airsoft-pelit*_{Y_SEM_ELOTON} ***eroavat*** *armeijasta*_{Z_SEM_ELOTON} *mm. siinä, että niissä keskitytään nimenomaan tuttujen ihmisten eliminointiin* (K854). Skeemojen välillä skeeman EROTA.ERI kytkeytyvät lisäksi argumenttien jäsennystrategiat ONP+*ja*+NP eli kategoria RAKENNE_KONNEKTORI (95.5 %), jota vastaa lause *Miten ***eroavat*** länsimainen pissi-kulttuuri ja lähi-idän/itä-afrikan pissis-kulttuuri?* (K959) sekä strategia O_{jaollinen} eli kategoria Y_JAOLLINEN (40.1 % skeemojen välillä; 12.1 skeeman sisällä), jota vastaa lause *Työskentelevät ***eroavat*** eri luonteisilla toisistaan* (K964). Vain skeemalla EROTA.PUOLISO on isompi painotus kategoriaan Y_JAOLLINEN (skeemojen välillä 50.8 %; skeeman sisällä 42.5 %). Reciprookki eli *toisistaan*-sanaa esiintyy skeemojen välillä lähes pelkästään (96.7 %, n=29) skeemassa EROTA.ERI. Havaintojen perusteella skeema EROTA.ERI manifestoi *erota*-lekseemin merkitysverkossa pikemminkin homonymiaa kuin muille *erota*-skeemoille läheistä sukua – polysemiaa tai merkityksen sumeutta (ks. luku 2.2.3) – olevaa semanttista suhdetta.

Skeema EROTA.IRTI esiintyy herkkyystason 3¹²⁶ merkitsevyydellä argumenttirakenteissa SV (37.4 %) – jota vastaa kategoria RAKENNE_VAIN_Y – ja SVA_{ero} (57 %) – jota vastaa kategoria RAKENNE_REKTIO. Agentti on ilmipantuna 80.3 %:ssa esiintymistä. Herkkyystasolla 3 skeemaan on kytköksissä yksilö- (62.5 %:a) ja instituutiotarkoitteiset agentit (25.0 %:a) ja herkkyystasolla 1 joukkotarkoitteiset agentit (12.5 %:a). Agentti on usein yleisesti tunnettu henkilö, kuten *Vanhanen*, jolloin lähteen ilmipaneminen ei ole välttämätöntä (Z-argumentti on poissa 36.8 %:ssa esiintymistä). Jos lähde on ilmipantuna (57.6 %), sen spesifioi 97.8 %:ssa korpusesimerkeistä luokkaan INSTITUUTIO koodaamani entiteetti, esimerkiksi *euro* tai *tehtävä*, jollainen esiintyy lauseessa *Suomen pitäisi **erota** eurosta* (K1047). Kun agentin spesifioi luokan Z_SEM_INSTITUUTIO tarkoite, kuten lauseessa *Kirkko **voisi erota valtiosta***, (K1051) kyse on itseasiassa usein skeemasta EROTA.IRTI2 (n=8, ks. s. 33), jonka olen yhdistänyt analyyseissä luokkaan EROTA.IRTI. Passiivia ei esiinny lainkaan.

Skeema EROTA.KIRKKO suosii argumenttirakennetta (S)VA_{ero} (82.7 %). Lähde on ilmipantu 81.9 %:ssa esimerkeistä, ja sen spesifioi lähes aina sana kategorian Y_SEM_INSTITUUTIO sana, nimittäin *kirkko*. Nämä ehdot täyttyvät 76.9 %:ssa (n=90) skeeman esiintymistä, kuten esimerkissä *Kirkolla ei pitäisi olla huolta, jos kirkosta Z_SEM_INSTITUUTIO **eroaa** sellaiset, jotka ei ole uskovaisia Y_SEM_JOUKKO*. Agentti – milloin ilmipantu (62.7 %) – esiintyy usein nolla-persoonassa (27.6 %) tai verbiin inkorporoituna (21.6 %). Inkorporoitu agentit painottuvat skeemojen välilläkin 47.2 %:a skeemaan EROTA.KIRKKO. Ilmipantu agentti on lähes yhtä usein joukko- (43.6 %) kuin yksilötarkoitteinen (56.4 %). Skeemalla ilmaistaankin toisaalta puhujan omia kirkosta eroamisista, toisaalta ihmisryhmien kirkosta eroamia vastakohtana skeemalle EROTA.IRTI, jonka agentteina esiintyy pääjohtajia, ministereitä, valtiota ja laitoksia. Skeema EROTA.KIRKKO luenta syntyykin ilmipannun rektioadverbiaalin avulla, kun taas skeeman EROTA.IRTI tarkoitteet sisältävät jo sinänsä vihjeen siitä, mistä yhteydestä he ehkä eroavat. Siksi sitä esiintyy useammin argumenttirakenteessa SV. Analyysin perusteella ei voida kuitenkaan väittää, että EROTA.IRTI ja EROTA.KIRKKO edustaisivat jyrkästi ”eri merkityksiä”. Kuten jo luvussa 2.2.3 aavistelin, voidaan ajatella, että niitä pikemminkin yhdistää skemaattisempi konstruktio ’erota jonkin yhteisön tms. yhteydestä’. Merkitys syntyy kontekstin vihjeiden avulla eli sanan merkitys joustaa kontekstin mukaan.

Skeema EROTA.PUOLISO hakeutuu herkkyystasolla 3 argumenttirakenteeseen S_{jaollinen}V, jota vastaa kategoria RAKENNE_Y_JAOLLINEN (42.5 %) sekä argumenttirakenteeseen V, jota vastaa kategoria RAKENNE_NA (23.3 %). Myös skeemojen välillä näiden kategorioiden havainnot keskittyvät skeemaan EROTA.PUOLISO (RAKENNE_Y_JAOLLINEN = 50.8 %;

¹²⁶ Eri taulukoiden herkkyystasojen mittaustulokset eivät ole keskenään suoraan vertailukelpoisia.

RAKENNE_NA=50.0). Ilmipantuina agentteina (60 %) esiintyykin kollektiivisanoja (42.5 %), kuten *lesbopari* tai *vanhemmat*. Lähde (Z-argumentti) on ilmipantuna 13.6 %:ssa (n=9) esiintymistä (eli 86.3 % esiintymistä sisältää pelkän agentin). Samalla se tarkoittaa luokan RAKENNE_REKTIO (SVA_{ero}) karttamista ja luokan z_MUOTO_NA (rektioadverbiaali ilmipane-matta) suosimista herkkyystasolla 3, joka on herkkyystasoista merkitsevin. Lähteen spesifioi aina luokan ELOLLINEN tarkoite eli yksilö tai joukko. Semanttis-syntaktisen käyttäytymisensä nojalla skeema EROTA.PUOLISO poikkeaa selvästi muista *erota*-skeemoista, esimerkiksi skee-masta EROTA.ERI, joka herkkyystasolla 3 hylkii luokkaa z_MUOTO_NA.

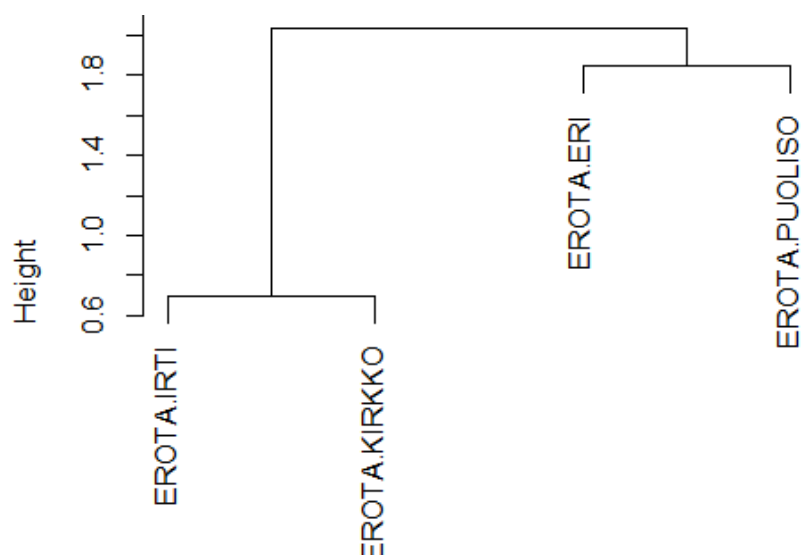
On hankala sanoa, sisältyykö *erota*-verbin ”intersubjektiiviseen merkitysverkkoon” erillinen ”solmu” (ks. Raukko 1994), jota vastaa sanakirjojen parafraasi ’erota parisuh-teesta’. Ilmeisesti sisältyy, sillä elämässä on ilmiö ’parisuhteen päättyminen’, johon suomen kieli viittaa lekseemillä *ero* (ERO.PUOLISO). Tuntuu epäuskottavalta, että puhuja joka kerta derivoisi *erota*-sanana *ero*-kantasemeemistä. *erota*-verbin luennat kuitenkin virittynevät Y- ja Z-argumenttien semanttisten vihjeiden nojalla – tai kontekstin laukaisemana. Sinclairin (1998: 6–8) mukaan lausuman merkitystä ei tule koostaa sanojen merkityksiä summaamalla vaan päinvastoin: leksikaalisen yksikön merkitys tulee koostaa tekstuaalisen ympäristön vih-jeistä ja merkityspiirteistä.

Tällaisessa näkemyksessä *erota*-verbin merkitys voidaan ymmärtää skemaattiseksi. Jos lähdettä ei ole ilmipantu, agentti laukaisee tulkinnan. Kun *kohumissi eroaa*, luenta on, että kyse on lemmensuhteen päättymisestä, mutta kun *ministeri eroaa*, luenta on, että kyse on ministerin työsuhteen päättymisestä ja tietyistä yhteiskunnallis-poliittisista uudelleenjär-jestelyistä. Lause *jalkapallo-ottelun katsominen eroaa* on omituinen, ellei siihen kytkeydy elatiivisijaista täydennystä, joka spesifioi että jalkapallo-ottelun katsominen eroaa esimer-kiksi *hevospoolo-ottelun katsomisesta*. Tämän sopii Firthin (1957: 11) sanat: “you shall know a word by the company it keeps”, kuitenkin sillä erotuksella, että kyse ei ole sanasta, vaan pikemminkin semeemistä (vrt. Jääskeläinen 2004: 47).

Sen sijaan ”indefiniittisen agentin” tokaisu *oli pakko erota* tai *ajattelinpa itsekkin erota* voi tarkoittaa ”mitä tahansa” riippuen siitä, kuka sen sanoo, kenelle, missä tilanteessa ja millä tavalla. Langacker (2008: 446) käyttää käsitettä *current discourse space* (CDS, va-paasti suomennettuna *jatkuvasti päivittyvä vuorovaikutustila*), joka tarkoittaa, että aiemmin sanotut tai kirjoitetut asiat vaikuttavat siihen, miten myöhemmin sanotut ja kirjoitetut asiat ymmärretään ja miten tekstin tai puheen odotetaan jatkuvan. Jos on puhe kirkosta, niin to-kaisija aikoo erota *kirkosta*, mutta jos on juuri ollut puhetta pienoislennokkikerhon jäsen-maksujen törkeästä nousemista, niin silloin erotaankin kai *pienoislennokkikerhosta*.

Skeeman EROTA.KIRKKO toteumat esiintyvät siksi harvoin ilman elatiivisijaista täydennystä, että harva entiteetti on sellainen, että maailmantiedon perusteella entiteetin ensisijainen eroamisen lähde kytkeytyisi kirkkoon tai uskonnolliseen yhteiskuntaan. Jos ajattelin erota *kirkosta*, on ”kristillinen tulkintakehys” jollain tavoin aktivoitava. Korpuksessani asiantilaa onkin näin: kun skeeman EROTA.KIRKKO toteuma esiintyy ilman Z-argumenttia, on kielenkäyttöympäristö aina sellainen, ettei tulkinta voi mennä metsään. Lähteen ilmoittamatta jättäminen skeeman EROTA.PUOLISO esiintymisissä on tarpeetonta samasta syystä: ko(n)teksti on aktivoinut *erota*-verbille parisuhteesta eroamisen tulkintahorisontin. Kollektiivitarkoitteiset agentit (*vanhemmat* ja *lesbopari*) taas ovat olentoja, jotka maailmantiedon nojalla erotessaan eroavat toisistaan, eivät esimerkiksi hallituksesta tai lennökkikerhosta.

Tarkastelen lopuksi klusterianalyysin visiota *erota*-verbistä. Klusteroitavana ovat *erota*-skeemat, joiden käyttäytymisprofileihin on syötetty yksimuuttujaisessa analyysissä mukana olleet luokat samankaltaisine luokkayhdistelyine kuin khiin neliö -testeissä¹²⁷.



Kuvio 4. *erota*-verbin klusterianalyysi. 27 ID-merkkiä. Etäisyys: *euclidean*, Metodi: *Ward*. Klusteristruktuurista hahmottuu rakenne, joka tukee jo tehtyjä havaintoja. Ensimmäisenä yhdistyy klusteri, johon kuuluvat skeemat EROTA.IRTI ja EROTA.KIRKKO. Aikainen yhdistyminen vihjaa, että ovat ne suhteellisen samanlaisia. Huomattavasti myöhemmin yhdistyy klusteri, johon kuuluvat EROTA.ERI ja EROTA.PUOLISO. Myöhäisestä yhdistymisestä voidaan päätellä, että niiden samankaltaisuus on verrattain vähäistä ja että yhdistyminen johtuu siitä, että klusterialgoritmin on aina pakko yhdistää skeemat johonkin klusteriin¹²⁸.

¹²⁷ Luokkayhdistelyt ovat lesummat kuin khiin neliö -testissä, ja mukana on myös pienifrekvenssisia luokkia. NA-luokat kategorioista Z_SEM ja Y_SEM on poistettu.

¹²⁸ Myös PAM-arvo (ks. luku 6.2) tukee päätelmää, sillä kahden klusterin ratkaisun luokittelutarkkuus on 25 %, kun taas kolmella klusterilla tarkkuus on 100 %.

En jatka pohdintaa tämän pitemmällä. *erota*-verbiin palaan luvussa 5.5 ja siirryn nyt *erottaa*-verbin analyysiin näistä tutkimustuloksista viisastuneena.

5.3 *erottaa*-verbi

erottaa-verbin tarkastelussa mukana Y-argumentin (patientti) ja Z-argumentin (lähde) lisäksi on X-argumentti, joka mittaa agenttia. Luvussa 5.3.1 tarkastelen muuttujia Y_SEM ja Z_SEM, luvussa 5.3.2 muuttujia Y_MUOTO ja Z_MUOTO ja luvussa 5.3.3 muuttujia X_SEM ja X_MUOTO, kunnes luvussa 5.3.4 tarkastelen argumenttirakennetta koskevia muuttujia RAKENNE ja RESIPROOKKI. Lopuksi luvuissa 5.3.5 ja 5.3.6 esitän yhteenvedot tuloksista.

X²-testien rajoitusten vuoksi ja myös tutkimuksen seurattavuuden takia yhdistelin skeemoja¹²⁹. Analyyseissä *erottaa*-verbin 9 skeemaa on yhdistetty 5 skeemaksi, jotka ovat EROTTAA.MUUT (← EROTTAA.ABSTRAKTI, EROTTAA.LUOKITTELU, EROTTAA.IRTI2 ja EROTTAA.KONKRETIA), EROTTAA.HAVAITA1, EROTTAA.HAVAITA2, EROTTAA.IRTI, EROTTAA.ERI. Aloitan patienttia tarkastelevasta muuttujista Y_SEM ja Y_MUOTO.

5.3.1 Muuttujat Y_SEM ja Y_MUOTO

Aloitan tarkastelemalla muuttujaa Y_SEM, jonka jakauma luokan NA (n=11) poistamisen jälkeen näkyy taulukossa 31. Taulukossa 31b on esillä muuttujan prosentuaaliset jakaumat.

Taulukko 31. Muuttujan Y_SEM jakauma (n=498). Suurimmat arvot on tummennettuina.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU	Σ _m
ELOTON	17	4	1	15	80	50	167
KONKRETIA	31	3	0	12	42	2	90
INSTITUUTIO	33	4	3	0	2	3	45
JOUKKO	19	14	41	0	18	5	97
YKSILO	6	8	73	1	11	0	99
Σ _s	106	33	118	28	153	60	498

Taulukko 31b. Muuttujan Y_SEM prosentuaalinen jakauma skeemojen sisällä (tumma tausta) ja skeemojen välillä (vaalea tausta).

	MUUT		ERI		IRTI		HAV1.		HAV2.		LUOKIT.		Σ _m
ELOTON	10.2	16.0	2.4	12.1	0.6	0.8	9.0	53.6	47.9	52.3	29.9	83.3	33.5
KONKRETIA	34.4	29.2	3.3	9.1	0.0	0.0	13.3	42.9	46.7	27.5	2.2	3.3	18.1
INSTITUUTIO	73.3	31.1	8.9	12.1	6.7	2.5	0.0	0.0	4.4	1.3	6.7	5.0	9.0
JOUKKO	19.6	17.9	14.4	42.4	42.3	34.7	0.0	0.0	18.6	11.8	5.2	8.3	19.5
YKSILO	6.1	5.7	8.1	24.2	73.7	61.9	1.0	3.6	11.1	7.2	0.0	0.0	19.9
Σ _s	21.3 %		6.6 %		23.7 %		5.6 %		30.7 %		12 %		100

¹²⁹ Skeemojen yhdistely pohjaa havaintoihini muuttujien X_SEM, Y_SEM, Z_SEM sekä RAKENNE jakaumista. On lukuisia tapoja yhdistää skeemoja, eikä ole selvää mikä niistä on paras. Suosin 6 skeeman ratkaisua suurimassa osassa *erottaa*-verbin analyysiä.

Silmäilemällä taulukkoja 31 ja 31b havaitaan, että isoinmalla luokalla ELOTON (n=167) on valta-asemassa skeemoissa EROTTAA.LUOKITTELU (83.3 %), EROTTAA.HAVAITA1 (53.6 %) ja EROTTAA.HAVAITA2 (52.3 %), joista kahteen viimeisimpään yhdistyy lisäksi luokka KONKRETIA (EROTTAA.HAVAITA1 = 42.9 %, EROTTAA.HAVAITA2 = 27.5 %). INSTITUUTIO:n esiintymät painottuvat skeemojen välillä (73.3 %) ja sisällä (31.1 %) skeemaan MUUT, ja sen sisällä skeemaan EROTTAA.IRTI2 (ks. liite 1). Luokan JOUKKO esiintymät hajautuvat skeemojen välille; isoin sisäinen osuus on skeemoissa EROTTAA.ERI (42.4 %) ja EROTTAA.IRTI (34.7 %). Luokka YKSIÖ kytkeytyy erityisesti skeemaan EROTTAA.IRTI (sisäisesti = 61.9 %, skeemojen välillä = 73.7 %).

Koska X^2 -testisuure ($X^2=427.627$, $df=20$, $p < 2.2e^{-16}$) on reilusti suurempi kuin khiin neliön kriittinen arvo, X^2 ($df=20$, $\alpha=0.05$) = 31.410, ja koska p ($p < 2.2e^{-16}$) alittaa selvästi riskitason ($\alpha = 0.05$), voidaan päätellä, että muuttujan Y_SEM ja skeemojen välillä on kohtalaisen voimakas kytkös ($V=0.463$). Luokat ennustavat noin neljäsosan ($U_{skeema|muuttuja}=0.272$, $\lambda_{skeema|muuttuja}=0.336$) skeemojen vaihtelusta, samoin kuin skeemat ($U_{muuttuja|skeema}=0.290$, $\lambda_{muuttuja|skeema}=0.296$) luokkien vaihtelusta. Tulos puoltaa tulkintaa, että patienttien vaihtelun taustalla on systemaattisuutta. Poikkeamien sijaintia, merkitsevyyttä ja suuntaa tutkii erotustaulukko taulukko 31c (solukohtaiset X^2 -kertymät ja Pearsonin jäämät, joiden pohjalta taulukon 31c merkitsevät solut määrittävät, ovat taulukoissa 31x–31xb, ks. liite 1).

Taulukko 31c. Muuttujan Y_SEM havaittujen ja odotettujen arvojen erotus, suunta ja merkitsevyys herkkyytasolla 1 (vaalea tausta) ja 3 (tumma tausta). Merkitsevät havainnot on tummennettuna.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-/18.5	-/7.1	-/38.6	+/5.6	+/28.7	+/29.9
KONKRETIA	+/11.8	-/3.0	-/21.3	+/6.9	+/14.3	-/8.8
INSTITUUTIO	+/23.4	+/1.0	-/7.7	-/2.5	-/11.8	-/2.4
JOUKKO	-/1.6	+/7.6	+/18.0	-/5.5	-/11.8	-/6.7
YKSIÖ	-/15.1	+/1.4	+/49.5	-/4.6	-/19.4	-/11.9

Taulukosta 31c havaitaan, että herkkyytason 4 tuloksia on kolme. Ensinnäkin skeema EROTTAA.IRTI on negatiivisesti kytkeytynyt luokkaan ELOTON (-38.6 esiintymää) ja positiivisesti luokkaan YKSIÖ (+49.5 esiintymää), joista jälkimmäinen on painoarvoltaan merkitsevin havainto ($X^2=104.6$, ks. liite 1). Lisäksi EROTTAA.IRTI on negatiivisessa kytköksessä luokkiin KONKRETIA (-21.1) ja INSTITUUTIO ja positiivisessa luokkaan JOUKKO (+18.2). Sen patienteiksi hakeutuu siis elollisia tarkoitteita. Kolmas herkkyytason 3 tulos on INSTITUUTIO:n yhteys yhdistelmäskeemaan MUUT (+23.4 esiintymää) (ja sen sisällä skeemaan EROTTAA.IRTI2.) Neljäs herkkyytason 3 havainto on skeeman EROTTAA.LUOKITTELU

kytkös luokkaan ELOTON. Erot tulevat näkyvimmäksi, kun luokat ELOTON ja KONKRETIA sekä JOUKKO ja YKSIÖ yhdistetään. Tätä ilmentää taulukko 31e.

Taulukko 31e. Muuttujan Y_SEM havaittujen ja odotettujen arvojen erotus, suunta ja merkitsevyys herkkyytasolla 1 (vaalea tausta) ja 3 (tumma tausta). Merkitsevät havainnot on tummennettuna.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-/6.7	-/10.0	-/59.9	+/12.6	+/43.0	+/21.0
INSTITUUTIO	+/23.4	+/1.0	-/7.7	-/2.5	-/11.8	-/2.4
ELOLLINEN	-/16.7	+/9.0	+/67.6	-/10.0	-/31.2	-/18.6

Taulukosta paljastuu tendenssi. Vain skeemat EROTTAA.ERI (+9 esiintymää) ja etenkin EROTTAA.IRTI (+67.6 esiintymää) hakeutuvat lauserakenteisiin, joissa patienttina on elollinen tarkoite. Skeemat EROTTAA.HAVAITA1 (+12.6 esiintymää), EROTTAA.LUOKITTELU (+21.0 esiintymää) ja etenkin EROTTAA.HAVAITA2 (+43.0 esiintymää) hakeutuvat lauserakenteisiin, joissa patientti on eloton. Yhdistelmäskeema MUUT karttaa lauserakenteita, joissa esiintyy elollinen patientti ja suosii luokkaan INSTITUUTIO koodattuja olioita (esimerkiksi *kirkkoa*).

Lisätietoa voidaan yrittää saada tarkastelemalla muuttujaa Y_MUOTO, jonka jakauma on kuitenkin ongelmallinen. Päädyin NA:n (n=11) poistamisen jälkeen tarkastelemaan sitä luokkien NP (\leftarrow PERS, n=33) ja PRON.RAK (\leftarrow RAKENNE, n=44 ja PRON, n=33) dikotomiana. Taulukko 32 on esillä sen jakauma (n=498, MUUT \leftarrow EROTTAA.LUOKITTELU).

Taulukko 32. Muuttujan Y_MUOTO jakauma.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	Σ_m
NP	145	32	117	19	108	421
PRON.RAK	21	1	1	9	45	77
Σ_s	166	33	118	28	153	498

Taulukko 32b. Muuttujan Y_MUOTO prosentuaalinen jakauma skeemojen sisällä (tumma tausta) ja skeemojen välillä (vaalea tausta).

	MUUT		ERI		IRTI		HAVAITA1		HAVAITA2		Σ_m
NP	34.4	87.3	7.6	97	27.8	99.2	4.5	67.9	25.7	70.6	84.5
PRON.RAK	27.3	12.7	1.3	3	1.3	0.8	11.7	31.1	58.4	29.4	15.5
Σ_s	33.3 %		6.6 %		23.7 %		5.6 %		30.7 %		100

Havaitaan, että luokan PRON.RAK – joka edustaa lähinnä anaforisia pronomineja ja lausetäydennyksiä – esiintymät jakautuvat skeemoihin EROTTAA.HAVAITA2 (n=45), EROTTAA.HAVAITA2 (n=9) ja MUUT (n=21), jossa isoin edustus on skeemoilla EROTTAA.ABSTRAKTI (n=4) ja EROTTAA.LUOKITTELU (n=13). Pronomini objektina on mahdollinen kaikissa skeemoissa, mutta lausetäydennyksiä (luokkaa RAKENNE) esiintyy vain näillä neljällä skeemalla. Skeemojen ja luokkien välillä on ehkä jonkinlaista systemaattista vaihtelua. Koska khiin neliön testisuure ($X^2=52.9252$, $df=4$, $p=8.835e^{-11}$) on suurempi kuin

kriittinen arvo X^2 ($df=4$, $\alpha=0.05$) = 9.489, ja koska p ($p=8.835e^{-1}$) alittaa riskitason ($\alpha = 0.05$), on näiden välillä todennäköinen myös tilastollinen kytkös, jonka voimakkuus on kohtalaista ($V=0.329$). Tällä kertaa havaitaan kuitenkin että P^{130} ei ole 1 vaan 0.9999967. Taulukko 32c esittää vielä lopuksi erotustaulukon, jossa tummalla pohjalla herkkyytason 1 merkitsevyyden määrittävät standardoidut Pearsonin jäämät.

Taulukko 32c. Muuttujan Y_MUOTO havaittujen ja odotettujen arvojen erotus, suunta ja merkitsevyys herkkyytastolla 1 (vaalea tausta) ja 3 (tumma tausta). Merkitsevät havainnot on tummennettuna.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2
NP	+ /4.7	+ / 4.1	+ / 17.2	- / 4.7	- / 21.3
PRON.RAK	- /4.7	- / 4.1	- / 17.2	+ / 4.7	+ / 21.3

Havaitaan, että yhdistelmäluokka PRON.RAK on positiivisessa assosiaatiossa erityisesti skeeman EROTTAA.HAVAITA2. Vastaavasti negatiivinen kytkös on merkitsevintä skeemassa EROTTAA.IRTI. Skeeman MUUT havainnot eivät ole merkitseviä. Havainnot tukevat jo tehtyjä löytöjä. Skeema EROTTAA.IRTI patienttina esiintyy lähinnä elollisia olioita, jotka lukeutuvat luokkiin NP (*Erotetaan ministeri*) ja sen sisällä luokkaan PERS (*Erotetaan hänet*). Skeeman EROTTAA.HAVAITA2 patienttina taas esiintyy elottomia tarkoituksia, jolloin luokat PRON ja RAKENNE tulevat kysymykseen. Jatkan tutkimalla rektioadverbiaalin kytköstä skeemoihin.

5.3.2 Muuttujat Z_SEM ja Z_MUOTO

X^2 -testin takia yhdistin muuttujan Z_SEM luokat ryhmiin ELOLLINEN (\leftarrow JOUKKO ja YKSILO), INSTITUUTIO, ELOTON (\leftarrow KONKRETIA). Luokan NA ($n=167$) poistamisesta seuraa taulukoiden 33-33b jakaumat. LUOKITTELU on jatkossa purettu erilleen yhdistelmäskeemasta MUUT.

Taulukko 33. Muuttujan Z_SEM jakauma ($n=342$). Isoimmat arvot on tummennettuina.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU	\sum_m
ELOTON	39	9	0	9	98	41	196
INSTITUUTIO	36	4	42	0	3	4	89
ELOLLINEN	16	19	0	0	19	3	57
\sum_s	91	32	42	9	120	48	342

¹³⁰ Cohen (1992) ehdottaa, että nollahypoteesin (valitulla merkittävyytastolla) kumoavan p -arvon lisäksi tuloksia esittäessä tulisi ilmoittaa *effektikoko* eli vapaasti suomennettuna *vaikuttavuus* (*Effect Size*, w) ja/tai voima (*Power*, $1-\beta$), sillä ”highly significant p -values should not be interpreted as automatically reflecting large effects” (Rosnow & Rosenthal 1989: 1279). β on todennäköisyys sille, että tilastollista merkitsevyyttä ei havaita, vaikka sitä ei ole, mikä vastaa tyypin II virhettä (Sullivan ja Feinn 2014). Effektikoon ja voiman ilmoittaminen on myös Arppen (2008) noudattama käytäntö – ainakin toisinaan. Olen ilmoittanut voiman (P) ainoastaan silloin, kun sen arvo on muu kuin 1. Cohenin (1992: 157) terminologiassa pieni vaikuttavuus (*a small effect size*, $w=0.10$) tarkoittaa, että todellisuudessa on ilmiö, jonka havaitseminen edellyttää hyvin perinpohjaista analyysiä. Suuri vaikuttavuus (*a large effect size*, $w=0.50$) tarkoittaa ilmiötä, joka on niin huomattava, että sen pystyy havaitsemaan paljaalla silmällä (*with the naked eye*). Tällainen on vaikkapa miesten ja naisten pituusero koh-talaisessa otoksessa. En ilmoita vaikuttavuuksia, mutta ne myötäilevät yleensä Cramérin V :n tuloksia.

Taulukko 33b. Muuttujan Z_SEM prosentuaalinen jakauma skeemojen sisällä (tumma tausta) ja skeemojen välillä (vaalea tausta).

	MUUT		ERI		IRTI		HAV1.		HAV2.		LUOKIT.		Σ_m
ELOTON	19.9	42.9	4.6	28.1	0.0	0.0	4.6	100	50.0	81.7	20.9	85.4	57.3
INSTITUUTIO	40.4	39.6	4.5	12.5	47.2	100	0.0	0.0	3.4	2.5	4.5	8.3	26.0
ELOLLINEN	28.1	17.6	33.3	59.4	0.0	0.0	0.0	0.0	33.3	15.8	5.3	6.2	16.7
Σ_s	26.6 %		9.4 %		12.3 %		2.6 %		35.1 %		14 %		100

Taulukkoja 33 ja 33b silmäilemällä havaitaan selvää systemaattisuutta. Ensinnäkin luokan INSTITUUTIO esiintymistä skeemassa EROTTAA.IRTI on skeemojen välisesti 47.2 % ja sisäisesti 100.0 %. Yksilöllinen on myös skeeman EROTTAA.HAVAITA1 sisäinen jakauma, josta 100 % (n=9) on luokassa ELOTON (ja sen sisällä kategoriassa KONKRETIA). Sillä on kuitenkin niin pieni kokonaisjakauma, ettei tuloksia voida varauksetta yleistää. Lisäksi luokalla ELOTON on valta-asema skeemoissa EROTTAA.LUOKITTELU (85.4 %), EROTTAA.HAVAITA2 (81.7 %) ja MUUT (42.9 %), kun taas skeemassa EROTTAA.IRTI sen osuus on 0 %. Luokka ELOLLINEN, jolla on valta-asema skeemassa EROTTAA.ERI (59.4 %), karttaa skeemoja EROTTAA.IRTI (0.0 %) ja EROTTAA.HAVAITA (0.0 %). Sitä esiintyy kohtalaisesti skeemoissa EROTTAA.HAVAITA2 (sisäisesti = 15.8 %; skeemojen välillä 33.3 %) ja EROTTAA.MUUT.

Koska testisuure X^2 ($df=10$, $p < 2.2e^{-16}$) = 231.395 ylittää khiin neliön kriittinen arvon X^2 ($df=10$, $\alpha=0.05$) = 18.307, ja koska p ($p < 2.2e^{-16}$) alittaa selvästi riskitason ($\alpha=0.05$), voidaan päätellä, että muuttujan Z_SEM ja skeemojen välillä on todennäköisesti melko voimakas kytkös ($V=0.582$). Skeemojen pohjalta voidaan ennustaa lähes kaksi viidesosaa ($U_{\text{muuttuja|skeema}}=0.352$, $\lambda_{\text{muuttuja|skeema}}=0.356$) luokkien käyttäytymisestä, kun taas luokat ennustavat noin viidesosan skeemojen käyttäytymisestä ($U_{\text{skeema|muuttuja}}=0.216$, $\lambda_{\text{skeema|muuttuja}}=0.176$). Skeemat ovat jälleen voimakkaammin kytköksissä luokkiin kuin luokat skeemoihin, mikä tarkoittaa, että tietämällä, mikä skeema, on parempi mahdollisuus tietää, mikä luokka kuin päinvastoin. Taulukko 33c tarkastelee merkitsevyyden sijaintia, suuntaa ja voimakkuutta (herkkyysarvot määrittyvät taulukon 33x pohjalta, ks. Liite 1).

Taulukko 33c. Muuttujan Z_SEM havaittujen ja odotettujen arvojen erotus, suunta ja merkitsevyys herkkyystasolla 1 (vaalealla taustalla) ja 3 (tummalla taustalla). Merkitsevät havainnot on tummennettuna.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-/13.2	-/9.3	-/24.1	+/3.8	+/29.2	+/13.5
INSTITUUTIO	+/12.3	-/4.3	+/31.1	-/2.3	-/28.2	-/8.5
ELOLLINEN	+/0.8	+/13.7	-/7.0	-/1.5	-/1.0	-/5.0

Taulukko 33c silmäilemällä havaitaan karkeasti kaksi ryhmää, joista toinen suosii lähteenä luokkaa ELOTON (EROTTAA.HAVAITA1, EROTTAA.HAVAITA2 ja EROTTAA.LUOKITTELU) ja toinen INSTITUUTIO:ta (EROTTAA.IRTI, EROTTAA.MUUT) tai luokkaa ELOLLINEN (EROTTAA.ERI).

Herkkyystasolla 3 havaintoja on neljä, joista painavin ($X^2 = +88.3$, ks. liite 1) on se, että skeema EROTTAA.IRTI esiintyy vain lauserakenteissa, joissa ilmipantu lähde kuuluu luokkaan INSTITUUTIO. Skeeman EROTTAA.HAVAITA2 lähteinä esiintyy odotettua useammin luokkaan ELOTON (+29.2) koodattuja tarkoitteita ja odotettua vähemmän INSTITUUTIO:ta (-28.2). Vahvistusta saa havainto, että skeema EROTTAA.HAVAITA1 kytkeytyy positiivisesti luokkaan ELOTON (KONKRETIA). Herkkyystasolla 3 merkitsevä on lisäksi skeeman EROTTAA.ERI kytkös luokkaan ELOLLINEN ($X^2=+35$, ks. liite 1). Toisin kuin EROTA.ERI se suosii siis lauserakenteita, joissa sekä patientin että lähteen spesifioi elollinen olio (esim. *Minut erottaa perus keskiluokkaisesta suomalaisista se, että iltakaljan sijaan nautin iltajointin*). Skeema EROTTAA.MUUT osoittaa positiivista kytköstä INSTITUUTIO:on ja negatiivista kytköstä luokkaan ELOTON (toisin kuin prosenttijakauma antaa olettaa, vrt. luku 5.5).

Tutkin lopuksi muuttujaa Z_MUOTO, jonka luokat järjestelin testausta varten näin: NA (\leftarrow NOLLA, n=2), NP (\leftarrow PERS, n=1) ja PRON.RAK (\leftarrow PRON, n=14 ja RAKENNE, n=22). Taulukko 34 esittää muuttujan jakauman ja 34b prosenttijakauman.

Taulukko 34. Muuttujan Z_MUOTO jakauma (n=509).

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	\sum_m
NA	29	2	82	19	35	167
NP	132	25	42	7	100	306
PRON.RAK	7	7	0	2	20	36
\sum_s	168	34	124	28	155	509

Taulukko 34b. Muuttujan Z_MUOTO prosentuaalinen jakauma skeemojen välillä (vaalealla taustalla) ja skeemojen sisällä (tummalla taustalla).

	MUUT		ERI		IRTI		HAVAITA1		HAVAITA2		\sum_m
NA	17.4	17.3	1.2	5.9	49.1	66.1	11.4	67.9	21.0	22.6	32.8
NP	43.1	78.6	8.2	73.5	13.7	33.9	2.3	25.0	32.7	64.5	50.1
PRON.RAK	19.4	4.2	19.4	20.6	0.0	0.0	5.6	7.1	55.6	12.9	7.1
\sum_s	33 %		6.7 %		24.4 %		5.5 %		30.5 %		100

Taulukoista havaintaan kaksi ryhmää, joista ensimmäinen on suosii ilmipantua lähdettä (EROTTAA.MUUT, EROTTAA.ERI ja EROTTAA.HAVAITA2) ja toinen on NA-myönteinen (EROTTAA.IRTI ja EROTTAA.HAVAITA1). Koska testisuure $X^2 (df=8, p < 2.2e^{-16}) = 135.269$ ylittää khiin neliön kriittinen arvon $X^2 (df=8, \alpha=0.05) = 15.507$, ja koska $p (p < 2.2e^{-16})$ alittaa selvästi riskitason ($\alpha=0.05$), voidaan päätellä, että muuttujan Z_MUOTO ja skeemojen välillä on todennäköisesti kohtalainen kytkös ($V=0.365$). Tuloksia voidaan syventää erotustaulukon 34c avulla (jonka herkkyystasot määrittyy liitteen 1 taulukosta 34x).

Taulukko 34c. Muuttujan *z_MUOTO* odotettujen ja havaittujen arvojen erotus. Herkkyystaso 3 (tumma tausta) ja herkkyystaso 1 (vaalea tausta). Merkitsevät havainnot on lihavoitu.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2
NA	-/26.1	-/9.2	+/41.3	+/9.8	-/15.9
NP	+/31.0	+/4.6	-/32.5	-/9.8	+/6.8
PRON.RAK	-/4.9	+/4.6	-/8.8	+/0.0	+/9.0

Taulukkoja silmäilemällä havaitaan, että luokka NA – joka tarkoittaa ettei rektioadverbiaalia ole lauserakenteessa – on positiivisessa kytköksessä skeemoihin EROTTAA.HAVAITA1 (+9.8 esiintymää) ja EROTTAA.IRTI. (+41.3 esiintymää), joista jälkimmäinen on ainoa herkkyystason 3 havainto ($X^2=+42$, ks. liite 1). Ne viihtyvät lauserakenteissa, joissa ilmipantuna on vain patientti ja ainoana skeemoina hylkivät luokkaa NP, johon on koodattu valtaosa ilmipannuista lähteistä (n=306). Luokkaan NP odotettua vahvemmin kytköksissä on yhdistelmäskeema MUUT (+31 esiintymää). Luokkaan PRON.RAK ovat positiivisessa kytköksessä skeemat EROTTAA.ERI (+4.6 esiintymää) ja EROTTAA.HAVAITA2 (+9 esiintymää). Luokkaa RAKENNE – joka tarkoittaa muun muassa lause *se + että*-lauseita – esiintyy rektioadverbiaalin paikalla vain skeemoilla EROTTAA.HAVAITA2 (n=17), EROTTAA.LUOKITTELU (n=4) JA EROTTAA.IRTI (n=1). Jatkan tarkastelua syventymällä muuttujiin RAKENNE ja RESIPROOKKI.

5.3.3 Muuttujat RAKENNE ja RESIPROOKKI

On syytä todeta, ettei muuttuja RAKENNE mittaa *erottaa*-verbin koko argumenttirakennetta, vaan osaa siitä (vrt. *erota*-verbin tarkastelu, jossa RAKENNE kattoi koko argumenttirakenteen). Osarakenteen eristäminen kokonaisuudesta luo ehkä keinotekoisin mittaustilanteen, jonka mahdollisiin ongelmiin palaan luvun lopussa. RAKENNE sisältää luokkien yhdistelyn ja NA (n=7) poistamisen jälkeen neljä luokkaa, jotka ovat VAIN_Y (n=111), REKTIO (\leftarrow VAIN_Z, n=2), Y_JAOLLINEN (n=46) ja KONNEKTORI (n=100). Jakauma on taulukoissa 35.

Taulukko 35. Muuttujan RAKENNE jakauma (n=502). Lihavoituna ovat isoimmat havainnot.

	ERI	IRTI	HAVAITA1	MUUT	HAVAITA2	Σ_m
REKTIO	29	42	8	85	81	245
VAIN_Y	1	77	17	6	10	111
Y_JAOLLINEN	1	0	2	20	23	46
KONNEKTORI	3	0	1	55	41	100
Σ_s	34	119	28	166	155	502

Taulukosta ilmenee joitakin tendenssejä. Kiinnostavaa on skeemojen EROTTAA.HAVAITA ja EROTTAA.MUUT toisiaan muistuttava jakauma. Tarkemman tarkastelun vuoksi yhdistelmäskeema MUUT sisälle ”pakatut” skeemat – EROTTAA.ABSTRAKTI, EROTTAA.LUOKITTELU, EROTTAA.IRTI2 ja EROTTAA.KONKRETIA – voidaan purkaa erilleen. En esitä muuttujan

RAKENNE havaittuja frekvenssejä ”purkamisen” jälkeen, mutta taulukot 35b ja 35b2 esittävät prosenttiosuudet skeemojen sisällä ja skeemojen välillä.

Taulukko 35b. Muuttujan RAKENNE prosentuaalinen jakauma skeemojen sisällä.

	IRTI	HAV1.	LUOKITTELU	IRTI2	ABSTR.	HAV2.	KONKR.	ERI	\sum_m
REKTIO	35.3	28.6	30.0	50.0	60.0	52.3	75.0	85.3	48.8
VAIN_Y	64.7	60.7	3.3	0.0	6.7	6.5	4.5	2.9	22.1
Y_JAOLLINEN	0.0	7.1	16.7	6.2	3.3	14.8	15.9	2.9	9.2
KONNEKTORI	0.0	3.6	50.0	42.8	30.0	26.5	4.5	8.8	19.9

Taulukko 35b2. Muuttujan RAKENNE prosentuaalinen jakauma skeemojen välillä.

	IRTI	HAV1.	LUOKITTELU	IRTI2	ABSTR.	HAV2.	KONKR.	ERI
REKTIO	17.1	3.3	7.3	6.5	7.3	33.1	13.5	11.8
VAIN_Y	69.4	15.3	1.8	0.0	1.8	9.0	1.8	0.9
Y_JAOLLINEN	0.0	4.3	21.7	4.3	2.2	50.0	15.2	2.2
KONNEKTORI	0.0	1.0	30.0	14.0	9.0	41.0	2.0	3.0
\sum_s	23.7	5.6	12	6.4	6	30.9	8.8	6.8

Taulukkoja silmäilemällä havaitaan karkeasti katsoen kolme ”kaveripiiriä”. Muuttujan RAKENNE alakategorioista valta-asema on REKTIO (n=245) kaikissa muissa paitsi skeemoissa EROTTAA.IRTI (REKTIO = 35.3 %) ja EROTTAA.HAVAITA1 (REKTIO = 28.6 %), joissa yleisin on luokka VAIN_Y (EROTTA.A.IRTI = 64.7 %, EROTTAA.HAVAITA1 = 60.7 %). Tyypillisintä niille on esiintyä ilman rektioadverbiaalia, joskin semmoinen on mahdollinenkin, mikä tarkoittaa, että ne suosivat argumenttirakennetta (S)VO. Luokan VAIN_Y esiintymät painottuvat skeemojen välilläkin skeemoihin EROTTAA.IRTI (69.4 %) ja EROTTAA.HAVAITA1 (15.3 %), tosin molemmat mainitut luokat ovat yleisempiä skeemassa EROTTAA.IRTI johtuen havaintojen isommasta kokonaisjakaumasta. Lisäksi ne karttavat luokkia Y_JAOLLINEN ja KONNEKTORI.

Loput skeemat suosivat lauserakenteita, joissa esiintyvät sekä Y- että Z-argumentti – agentti ja lähde – eli luokkia REKTIO ja KONNEKTORI sekä luokkaa Y_JAOLLINEN, jossa Z-argumentti sisältyy Y-argumentin jaolliseen tai kollektiiviseen merkitykseen. Toiseen piiriin kuuluvatkin skeemat EROTTAA.LUOKITTELU, EROTTAA.IRTI2, EROTTAA.ABSTRAKTI sekä skeema EROTTAA.HAVAITA2, jonka ”paras kaveri” lieene EROTTAA.ABSTRAKTI. Tätä porukkaa yhteen liittää REKTIO, VAIN_Y ja etenkin KONNEKTORI, jonka esiintymät painottuvat skeemojen välilläkin kyseisiin skeemoihin. Sisäisesti suurin edustus KONNEKTORI:lla on skeemassa EROTTAA.LUOKITTELU (50 %). Kolmannen porukan muodostavat skeemat EROTTAA.KONKRETTIA ja EROTTAA.ERI luokkien KONNEKTORI ja VAIN_Y sekä eritoten luokan REKTIO perusteella (EROTTA.A.KONKRETTIA = 75.3 %, EROTTAA.ERI=85.3 %). Ne siis mielellään esiintyvät argumenttirakenteessa (S)VOA_{ero}. Ryhmien erot ovat kuitenkin pieniä, ja esimerkiksi luokka Y_JAOLLINEN (sekä osin luokka VAIN_Y) yhdistää skeemoja lähes kautta

linjan. Toisaalta kaveripiiriä kolme (EROTTA.KONKRETIA ja EROTTAA.ERI) yhdistää luokien VAIN_Y ja KONNEKTORI pieni yhteisosuus. Eikä olisi rikos liittää skeemaa EROTTAA.KONKRETIA kaveripiiriin kaksi ja samalla syöstä skeema EROTTAA.LUOKITTELU kaveripiiristä ulos. Luokan Y_JAOLLINEN prosentuaalinen jakauma on suhteessa pienin liikkuen välillä 0–16.7 % keskiarvon ollessa 8.9 %. Tasaisesta jakaumasta voidaan tehdä arvaus, ettei sillä ole skeemojen vaihtelussa merkitsevää roolia¹³¹.

X^2 -testiä varten yhdistin luokan Y_JAOLLINEN luokkaan KONNEKTORI¹³². Koska testitulos X^2 ($df=14$, $p < 2.2e^{-16}$) = 277.713 ylittää khiin neliön kriittinen arvon X^2 ($df=14$, $\alpha=0.05$) = 23.685, ja koska p ($p < 2.2e^{-16}$) alittaa selvästi riskitason ($\alpha=0.05$), voidaan päätellä, että muuttujan RAKENNE ja skeemojen välillä on todennäköisesti melko voimakas kytkös ($V=0.526$). Skeemojen pohjalta voidaan ennustaa noin neljäsosa ($U_{\text{muuttuja|skeema}}=0.277$, $\lambda_{\text{muuttuja|skeema}}=0.257$) luokkien käyttäytymisestä, kun taas luokat ennustavat alle viidesosan skeemojen käyttäytymisestä ($U_{\text{skeema|muuttuja}}=0.156$, $\lambda_{\text{skeema|muuttuja}}=0.193$). Tietämällä skeeman voimme siis jälleen paremmin ennustaa luokan kuin päinvastoin.

Erojen sijaintia, suuntaa ja merkitsevyyttä selvitän erotustaululla 35c (jonka merkitsevyyshavainnot määrittyvät liitteessä 1 olevan taulukon 35x pohjalta).

Taulukko 35c. Muuttujan RAKENNE poikkeamat. Herkkyystaso 3 on tummalla taustalla ja herkkyystaso 1 vaalealla taustalla. Merkitsevät solut on lihavoitu.

	IRTI	HAV1.	LUOKIT.	IRTI2	ABSTR.	HAV2.	KONKR.	ERI
REKTIO	-16.1	-5.7	-11.3	+0.4	+3.4	+5.4	+11.5	+12.4
VAIN_Y	+50.7	+10.8	-11.3	-7.1	-4.6	-24.3	-7.7	-6.5
KONNEKTORI	-34.6	-5.1	+22.5	+6.7	+1.3	+18.9	-3.8	-5.9

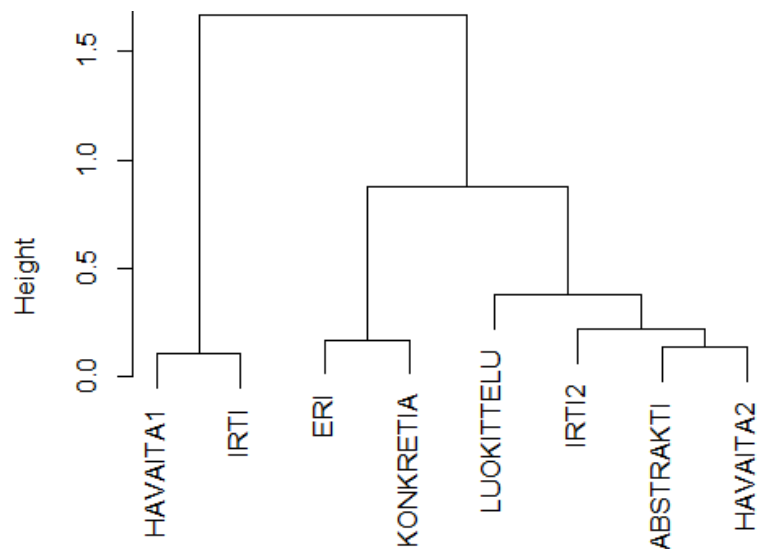
Herkkyystasolla 3 merkitseviä kytköksiä on kolme, joista painavimmat ovat skeeman EROTTAA.IRTI yliedustus (+50.7 esiintymää) luokassa VAIN_Y ($X^2=95.6$, ks. liite 1) ja aliedustus (-34.6 esiintymää) luokassa KONNEKTORI ($X^2=34.2$, ks. liite 1). Herkkyystason 3 havainto on skeeman EROTTAA.LUOKITTELU yliedustus (+22.5 esiintymää) luokassa KONNEKTORI ($X^2=29.1$, ks. liite 1). Herkkyystasolla 1 näkyy kiinnostavia ryhmiä. Skeemoja EROTTAA.KONKRETIA ja EROTTAA.ERI yhdistää luokan REKTIO suosiminen (yht. +23.9 esiintymää) ja luokan VAIN_Y (yht. -14.2 esiintymää) karttelu; vastakohtana toimii skeemojen EROTTAA.IRTI, ja EROTTAA.HAVAITA1 positiivinen kytkös luokkaan REKTIO. Luokkia EROTTAA.LUOKITTELU, EROTTAA.HAVAITA2 ja EROTTAA.IRTI2 yhdistää luokan KONNEKTORI

¹³¹ Tulosta tukee X^2 -jakaumien vertailu. Ne saadaan ristiintaulukoimalla skeemat suhteessa kunkin luokan läsnäoloon / poissaoloon. Tulokset ovat seuraavat: VAIN_Y ($X^2=207.877$, $p=2.2e^{-16}$), KONNEKTORI ($X^2=94.904$, $p=2.2e^{-16}$), REKTIO ($X^2=50.338$, $p=1.239e^{-08}$) ja Y_JAOLLINEN ($X^2=24.836$, $p=0.0008114$). Kaikki ylittävät khiin neliön kriittisen arvon X^2 ($df=7$, $\alpha=0.05$) = 14.067. Luokka Y_JAOLLINEN ei kuitenkaan läpäise testin ehtoja, sillä siinä odotettuja arvoja alle 5 on 31.25 % luokista.

¹³² Semanttisesti sen olisi voinut yhdistää myös luokkaan REKTIO.

yliedustus (yht. +48.1 esiintymää) ja luokan VAIN_Y aliedustus (yht. -42.7 esiintymää). Nämä yhdessä skeeman EROTTAA.ABSTRAKTI kanssa muodostavat jälleen kaveripiirin; yhteisestä linjasta poikkeaa lähinnä skeeman EROTTAA.LUOKITTELU negatiivinen suhde REKTIO:on. Tuloksista voidaan tehdä arvaus, että luokka EROTTAA.ABSTRAKTI ehkä edustaa skeemojen prototyyppiä argumenttirakenteen tasolla, koska sillä on kaikissa luokissa suurin piirtein odotettu määrä havaintoja.

Varaslähtönä lukuun 6 demonstroin lopuksi, kuinka hiarkkista klusterianalyysia voidaan käyttää apuvälineenä yksittäisenkin muuttujan analyysissä, vaikkei sitä siihen olekaan luotu. Koska klusterianalyysi ei aseta tarkkoja minimiehtoja luokkien koolle, käyttäytymisprofiileihin on syötetty edelliset 8 skeemaa, jotka on klusteroitu suhteessa muuttujan RAKENNE luokkiin REKTIO, VAIN_Y, KONNEKTORI, Y_JAOLLINEN ja NA, jolloin $n=509$.



Kuvio 5. Muuttuja RAKENNE. Etäisyys: *euclidean*, Metodi: *Ward* (ks. luku 6).

Havaitaan, että skeemat EROTTAA.IRTI ja HAVAITA1 muodostavat oman klusterin. Toisen klusterin muodostavat EROTTAA.ERI ja EROTTAA.KONKRETIA. Kolmannessa klusterissa on kolme lohkoa. Ensimmäisenä yhdistyvät EROTTAA.ABSTRAKTI ja EROTTAA.HAVAITA2, joihin yhdistyy EROTTAA.IRTI2 ja lopulta EROTTAA.LUOKITTELU. Myöhempi yhdistyminen viittaa, että samankaltaisuus vähenee, joskin se on yhä suhteellisen suurta verrattuna esimerkiksi klusteriin, johon kuuluvat EROTTAA.HAVAITA1 ja EROTTAA.IRTI. Klusteristruktuuri tukee edellä todettuja päätelmiä.

Ennen *erottaa*-verbin agenttia tarkastelen vielä muuttujaa RESIPROOKKI, jonka eristäminen *erottaa*-verbin argumenttirakenteen tarkastelusta saattaa olla yhtä keinoitekoista kuin X-argumentin eli agentin. Tutkin muuttujaa RESIPROOKKI taulukoiden 36, 36b ja 37 avulla. Taulukko 36 esittää muuttujan RESIPROOKKI jakauman ja 36b prosenttijakauman (LOPUT ←

EROTTAA.ERI, EROTTAA.IRTI, EROTTAA.HAVAITA1). Taulukko 37 esittää muuttujan RESIPROOKKI ristiintaulukoinnin.

Taulukko 36. Muuttujan RESIPROOKKI jakauma (n=509).

	IRTI2	LUOKIT.	HAV2.	KONKR.	ABSTR.	ERI	IRTI	HAV1.	\sum_m
NA	17	40	125	41	30	34	124	28	439
LÄSNÄ	15	20	30	3	2	0	0	0	70
\sum_s	32	60	155	44	32	34	124	28	509

Taulukko 36b. Muuttujan RESIPROOKKI prosentuaalinen jakauma skeemojen sisällä (tumma tausta) ja skeemojen välillä (vaalea tausta).

	IRTI2		LUOKIT.		HAV2.		KONKR.		ABSTR		LOPUT	
NA	3.9	53.1	9.1	66.7	28.5	80.6	9.3	93.2	6.8	93.8	42.4	100
LÄSNÄ	21.4	46.9	28.6	33.3	42.9	19.4	4.3	6.8	2.9	6.2	0.0	0

Taulukko 37. Muuttujan RAKENNE ja RESIPROOKKI ristiintaulukointi.

	REKTIO	NA	VAIN_Y	Y_JAOLLINEN	KONNEKTORI
NA	245	7	111	25	51
LÄSNÄ	0	0	0	21	49

Taulukosta 37 ilmenee sama tendenssi kuin *erota*-verbin kanssa: muuttujaa RESIPROOKKI esiintyy ainoastaan luokkien Y_JAOLLINEN ja KONNEKTORI kanssa – siis argumenttirakenteissa, joiden semanttinen valenssi sisältää kaksi täydennystä, joita ei spesifoida rektiotäydennyksellä. Tällaisiin lauserakenteisiin hakeutuvat skeemat EROTTAA.HAVAITA2 (n=30), EROTTAA.LUOKITTELU (N=20) ja etenkin EROTTAA.IRTI2 (n=15), jolla RESIPROOKKI:a esiintyy 46.9 %:ssa. Koska testisuure X^2 ($df=4$, $p < 2.2e^{-16}$) = 202.090 ylittää khiin neliön kriittisen arvon X^2 ($df=4$, $\alpha=0.05$) = 9.488, ja koska p ($p < 2.2e^{-16}$) alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että muuttujien RAKENNE ja RESIPROOKKI välillä on todennäköisesti melko voimakas kytkös ($V=0.630$)¹³³. Tietämällä argumenttirakenteen voimme joka toisessa tapauksessa ($U_{\text{resiprookki|rakenne}} = 0.504$) ennustaa, esiintyykö *toisistaan* lauserakenteessa vai ei. Toiseen suuntaan kytkös on heikompaa ($U_{\text{rakenne|resiprookki}}=0.158$).

Olen epäillyt, että saattaa olla keinotekoista tarkastella erikseen transitiiviverbin yksösargumenttia (agenttia, X-argumentti), kakkosargumenttia (patientti, Y-argumentti) ja kolmosargumenttia (lähde, Z-argumenttia) sekä RESIPROOKKI:a. Asian selvittämiseen voimme ristiintaulukoida muuttuja X_MUOTO (jonka luokkia ovat NP, NA, NOLLA) suhteessa muuttujaan RAKENNE ja RESIPROOKKI, ja suorittaa taulukoille khiin neliö -testit.

Muuttujien RAKENNE ja X_MUOTO ristiintaulukoinnin tulos X^2 -testissä on: X^2 ($df=6$, $p=0.0782$) = 11.348. Lisäksi P (power) = 0.709. Koska arvo on pienempi kuin khiin neliön

¹³³ Yhdessä solussa (NA:LÄSNÄ) odotettu arvo on 0.096, joten tuloksiin on suhtauduttava varauksella, joskin testin suorittaminen ilman luokka NA antaa lähes identtiset tulokset.

kriittinen arvo $X^2 (df=6, a=0.05) = 12.592$, ja koska $p (p=0.0782)$ on suurempi kuin riskitaso ($a=0.05$), nollahypoteesia ei voida hylätä, vaan se jää voimaan. Muuttujien RESIPROOKKI ja X_MUOTO tulos on: $X^2 (df=6, p=0.409) = 1.788$. Lisäksi P on vain 0.205. Tulkinta on, että muuttujien välillä ei ole tilastollisesti merkitsevää yhteyttä. Muuttujat ovat toisistaan riippumattomia ja ehkä ”vapaassa vaihtelussa”. Riippumattomuus tarkoittaa toisin sanoen, että RESIPROOKKI (eli *toisistaan*-lekseemi) ja muuttujan RAKENNE spesifioimat eri argumenttira-kennetyypit esiintyvät *erottaa*-verbin kanssa riippumatta siitä, saako muuttuja X_MUOTO arvokseen NP, NA vai NOLLA. X^2 -testillä ei kuitenkaan voida tarkastella kolmen tai useamman luokan välistä riippuvuutta¹³⁴. Tulos ei varmista, etteikö muuttujilla RAKENNE, RESIPROOKKI ja X_MUOTO voisi olla sidoksisuutta, joka ilmenisi esimerkiksi jonkin muun muuttujan yhteisvaikutuksesta. Nämä kysymykset ovat kuitenkin tämän tutkimuksen ulottumattomissa.

Siirryn seuraavaksi tarkastelemaan *erottaa*-verbin agenttia.

5.3.4 Muuttujat X_SEM ja X_MUOTO

erottaa-verbin agentin ominaisuuksia mittaavat muuttujat ovat X_SEM ja X_MUOTO, joista edellinen mittaa agentin semanttista ontologiaa ja jälkimmäinen agenttikonstituentin läsnäolon muotoa. Muuttujan jakaumaa taulukkoon 38 ja prosenttiluokkiin 38b havainnollistaakseni yhdistelin luokkia seuraavasti: NOLLA (nollapersoonaa), NA ja PRON (\leftarrow PRON, $n=19$ ja RAKENNE, $n=19$), PERS (\leftarrow PERS, $n=17$ ja INKORPOROITU, $n=26$).

Taulukko 38. Muuttujan X_MUOTO jakauma ($n=509$). Isoimmat havainnot lihavoituna.

	ERI	HAV2.	HAV1.	LUOKITTELU	IRTI	KONKR.	ABSTRAK.	IRTI2	\sum_m
NP	31	98	16	24	38	13	6	0	226
NA	3	12	0	11	79	31	23	32	191
NOLLA	0	45	12	25	7	0	3	0	92
\sum_s	34	155	28	60	124	44	32	32	509

Taulukko 38b. Muuttujan X_MUOTO prosentuaalinen jakauma skeemojen sisällä.

	ERI	HAV2.	HAV1.	LUOKITTELU	IRTI	KONKR.	ABSTRAK.	IRTI2	\sum_m
NP	91.2	63.2	57.1	40.0	30.6	29.5	18.8	0	44.4
NA	8.8	7.7	0.0	18.3	63.7	70.5	71.9	100	37.5
NOLLA	0.0	29.0	42.9	41.7	5.6	0.0	9.4	0	18.1

Taulukko 38b2. Muuttujan X_MUOTO prosentuaalinen jakauma skeemojen välillä.

	ERI	HAV2.	HAV1.	LUOKITTELU	IRTI	KONKR.	ABSTRAK.	IRTI2
NP	13.7	43.4	7.1	10.6	16.8	5.8	2.7	0.0
NA	1.6	6.3	0.0	5.8	41.4	16.2	12.0	16.8
NOLLA	0.0	48.9	13.0	27.2	7.6	0.0	3.3	0.0
\sum_s	6.7	30.5	5.5	11.8	24.4	8.6	6.3	100.0

¹³⁴ Tähän tarvittaisiin esimerkiksi regressioanalyysia, joka ei sisälly työni välineistöön.

Silmiinpistävin havainto on, että *erottaa*-verbin 283:sta korpuksimerkissä 509:stä (55.6 %) ei esiinny kooditussysteemi valossa ilmikoodattua agenttikonstituenttia¹³⁵. Agentin koodaamattomuuden perusteella erottuu kaksi ryhmää, joista ensimmäinen suosii luokkaa NOLLA (EROTTAA.HAVAITA1, 42.9 %; EROTTAA.LUOKITTELU, 41.7 % ja EROTTAA.HAVAITA2, 29 %) ja toinen NA:ta (EROTTAA.IRTI2, 100 %; EROTTAA.ABSTRAKTI, 71.9 %; EROTTAA.KONKRETIA, 70.5 % ja EROTTAA.IRTI, 63.7 %). Täysin ”agenttikammoinen” on EROTTAA.IRTI2. Jonkin verran ilmikoodattuja agenteja on skeemoilla EROTTAA.ABSTRAKTI (n=6), EROTTAA.KONKRETIA (n=13) ja EROTTAA.IRTI (n=38). Valta-asemassa NP – ilmikoodattu agentti – on kolmessa skeemassa, jotka ovat EROTTAA.ERI (91.2 %), EROTTAA.HAVAITA2 (63.2 %) ja EROTTAA.HAVAITA1 (57.1 %). Niillä on myös isoimmat yhteenlasketut agenttiosuudet (NOLLA + NP; EROTTAA.HAVAITA1 = 100 %, EROTTAA.HAVAITA2 = 92.2 % ja EROTTAA.ERI 91.2 % josta NOLLA = 0 %). Tämä vihjaa, että juuri niiden merkitysrakenne edellyttää transitiivisen tekoketjun ensimmäisenä jäsenenä toimivan kokijan tai aistijan (EROTTAA.HAVAITA1–2) tai muun (EROTTAA.ERI) rooliargumentin läsnäoloa.

Khiin neliö -testausta varten ”puran” NP-luokan kolmeen osaan, jotka ovat NP, PRON (PRON ja RAKENNE) ja PERS (PERS ja INKORPOROITU). Skeemojen lukumäärää on samalla supistettava siten, että NA:ta suosivaan yhdistelmäskeemaan MUUT ”pakataan” skeemat EROTTAA.IRTI2, EROTTAA.KONKRETIA ja EROTTAA.ABSTRAKTI. Tästä syntyy taulukko, jossa skeemojen lukumäärä on 6 ja luokkien 5 (jolloin soluja $> 5 = 16.7$ %).

Koska mainitun taulukon testisuure X^2 ($df=20$, $p < 2.2e^{-16}$) = 403.128 on merkittävästi suurempi kuin khiin neliön kriittinen arvo X^2 ($df=20$, $\alpha=0.05$) = 31.410, ja koska myös p ($p < 2.2e^{-16}$) on selvästi pienempi kuin riskitaso ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujan välillä lähes melko voimakas kytkös ($V = 0.445$). Assosiaatio on vähän vahvempaa skeemoista luokkia kohtaan ($U_{\text{muuttuja|skeema}}=0.245$, $\lambda_{\text{muuttuja|skeema}}=0.299$) kuin luokista skeemoja kohtaan ($U_{\text{skeema|muuttuja}}=0.213$, $\lambda_{\text{skeema|muuttuja}}=0.237$).

Ristiintaulukoimalla skeemat ja muuttujan X_MUOTO luokat yksitellen loppuja luokkia vasten voidaan selvittää, mikä luokista erottelee merkitystä parhaiten. Tulokseksi X^2 -testistä saadaan: NA = 214.77 ($df=5$, $p < 2.2e^{-16}$), PRON = 162.695 ($p < 2.2e^{-16}$), NOLLA = 158.249 ($p < 2.2e^{-16}$), PERS = 33.611 ($p=2.845e^{-06}$), NP = 23.589 ($p=0.00026$). Kaikki tulokset ylittävät khiin neliön kriittisen tason X^2 ($df=7$ $\alpha=0.05$) = 11.071, ja myös p -arvot ovat kriittistä merkitsevyystasoa alempia. Tulkinta on, että NA erottelee merkitystä parhaiten. Luokka NP on

¹³⁵ Jakauma olisi vieläkin isompi, jos luokka INKORPOROITU (n=26) olisi yhdistetty luokkaan NA. En näin menetellyt näin, sillä lause *Erotatko nurmikolla pökäleen* manifestoi agentin verbin persoonapäätteen muodossa. NA-luokasta 114 selittyy sillä, että lause on passiivissa, mutta 77 on aktiivilauseissa, jotka käsittävät 47 nesesiiivistä verbiketjua, 2 kysymyslausetta ja 28 infinitiivisubjektia tai muuta subjektitonta konstruktia.

jakautunut tasaisemmin, minkä vuoksi sen ja skeemojen yhteys on heikompaa.

Tarkastelen lopuksi vielä poikkeavuuksien sijaintia, suuntaa ja merkitsevyyttä taulukon 38c avulla (jonka herkkyytaset on laskettu liitteen 1 taulukosta 38x).

Taulukko 38c. Muuttujan X_MUOTO odotettujen ja havaittujen arvojen erotus. Herkkyytaso 3 (tumma tausta) ja herkkyytaso 1 (vaalea tausta). Merkitsevät havainnot on lihavoitu.

	MUUT	IRTI	ERI	HAVAITA1	HAVAITA2	LUOKITTELU
NA	+48.7	+32.5	-9.8	-10.5	-46.2	-11.5
NOLLA	-16.5	-15.4	-6.1	+6.9	+17.0	+14.2
PERS	-7.1	-6.5	-2.9	+4.6	+10.9	+0.9
PRON	-5.6	-6.6	+16.9	-0.7	-0.4	-3.7
NP	-16.3	-4.0	+1.8	-0.4	+18.7	+0.1

Taulukkojen tulokset tukevat jo tehtyjä löydöksiä. Herkkyytason 3 löytöjä on 3, joista painavin ($X^2 = 138.4$, ks. liite 1) on skeeman EROTTAA.ERI yhteys luokkaan PRON (+16.9 esiintymää). Tämä yhteys selittyy sillä, että skeeman EROTTAA.ERI korpustoteutumisissa subjektin paikalla on yleensä *se*-tukipronomini + *että* -rakenne. Muut herkkyytason 3 löydöt liittyvät luokkaan NA, jota skeema EROTTAA.MUUT suosii (+48.7 esiintymää) ja skeeman EROTTAA.HAVAITA2 hylkii (-46.2 esiintymää, $X^2 = -36.6$, ks. liite 1). Skeemojen välille on oppositiota myös suhteessa luokkiin NOLLA, PERS ja NP. Lauserakenteisiin, joissa agentti on ilmipantu (PERS, PRON ja NP), hakeutuvat skeemat EROTTAA.HAVAITA2 (yht.+29.6) ja EROTTAA.ERI (yht. +18.7). Nollapersoonaan positiivisesti assosioituvat skeemat EROTTAA.HAVAITA1, EROTTAA.HAVAITA2 ja EROTTAA.LUOKITTELU. Kun tarkkaillaan herkkyytason 1 löytöjen kokonaissuhdetta, havaitaan että skeemat muodostavat kaksi ryhmää. Ensimmäiseen kuuluvat EROTTAA.MUUT ja EROTTAA.IRTI, toiseen EROTTAA.HAVAITA1, EROTTAA.HAVAITA2 ja EROTTAA.LUOKITTELU; skeeman EROTTAA.ERI profiili on yksilöllisin.

Muuttujan X_SEM – joka tarkastelee *erottaa*-verbin agenttia – X^2 -testauksen ongelma on NA-luokan jättiedustus ($n=283$). Sen poistamisen jälkeen jäljelle jää 226 havaintoja, jotka jakautuvat epätasaisesti. Taulukkoa 39 varten luokkia on yhdistettävä esimerkiksi siten, että jäljelle jäävät ELOLLINEN (\leftarrow JOUKKO ja YKSIÖ), ELOTON (\leftarrow KONKRETIA) ja INSTITUUTIO. Skeemoina toimivat edelliset 6 kategoriaa. Taulukko 39b esittää prosenttijakaumat.

Taulukko 39. Muuttujan X_SEM jakauma ($n=226$).

	ERI	MUUT	HAVAITA1	HAVAITA2	LUOKITTELU	IRTI	\sum_m
ELOTON	31	8	3	3	0	0	45
INSTITUUTIO	0	3	1	6	2	19	31
ELOLLINEN	0	8	12	89	22	19	150
\sum_s	31	19	16	98	24	38	226

Taulukko 39b. Muuttujan X_SEM prosentuaalinen jakauma skeemojen välillä (vaalealla taustalla) ja skeemojen sisällä (tummalla taustalla).

	ERI		MUUT		HAVAITA1		HAVAITA2		LUOKIT.		IRTI		Σ_m
ELOTON	68.9	100	17.8	42.1	6.7	18.8	6.7	3.1	0.0	0.0	0.0	0	19.9 %
INSTITUUTIO	0.0	0.0	9.7	15.8	3.2	6.2	19.4	6.1	6.5	8.3	61.3	50	13.7 %
ELOLLINEN	0.0	0.0	5.3	42.1	8.0	75.0	59.3	90.8	14.7	91.7	12.7	50	66.4 %
Σ_s	13.7 %		8.4 %		7.1 %		43.4 %		10.6 %		16.8 %		100 %

Taulukoista 39 ja 39b erottuu oikeastaan neljä tyyppiä. Ensimmäinen, johon kuuluvat EROTTAA.LUOKITTELU (91.7 %), EROTTAA.HAVAITA2 (90.8 %) ja EROTTAA.HAVAITA1 (75 %), suosii elollisia agentteja. Toisessa (EROTTAA.ERI) esiintyy vain elottomia agentteja. Kolmanteen tyyppiin kuulu yhdistelmäskeema MUUT, jonka ilmipannut agentit jakautuvat ELOTTOMIIN (42.1 %), ELOLLISIIN (42.1 %) ja luokkaan INSTITUUTIO (15.8 %). Neljännessä tyypissä, johon kuuluu EROTTAA.IRTI agentit jakautuvat tasan elottomien ja instituution kesken. Esiintymien niukkuuden takia muuttujan X_SEM X^2 -testaus ei ole mahdollista ilman isoja luokkayhdistelyjä. Suoritin testin yhdistämällä luokat ELOLLINEN ja INSTITUUTIO, jolloin soluja, jotka ovat > 5 , oli edelleen 25 %, mikä tarkoittaa, ettei testaus täytä miniehtoja. Vaikka tulokset vaikuttaisivat lingvistisesti päteviltä, tilastotieteellisestä näkökulmasta niihin on suhtauduttava varauksella. Koska testisuuri, X^2 ($df=5$, $p < 2.2e^{-16}$) = 163.434, ylittää khiin neliön kriittinen arvon X^2 ($df=5$, $\alpha=0.05$) = 11.071, ja koska p ($p < 2.2e^{-16}$) alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujien välillä on todennäköisesti voimakas kytkös ($V=0.850$). Taulukko 39c esittää lopuksi poikkeavuuksien suunnan, merkitsevyyden ja havaittujen ja odotettujen arvojen erotuksen.

Taulukko 39c. Muuttujan X_SEM odotettujen ja havaittujen arvojen erotus, poikkeavuuden suunta ja merkitsevyys herkkyytasolla 3 (tumma tausta) ja 1 (vaalea tausta). Tilastollisesti merkitsevät havainnot on lihavoitu.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	+ / 4.2	+ / 24.8	- / 7.6	- / 0.2	- / 16.5	- / 4.8
ELOLLINEN	- / 4.2	- / 24.8	+ / 7.6	+ / 0.2	+ / 16.5	+ / 4.8

Skeemat jakautuvat tulosten valossa kolmeen ryhmään, jossa ensimmäisessä esiintyy elottomia agentteja (EROTTAA.MUUT ja etenkin EROTTAA.ERI) ja toisessa elollisia (EROTTAA.IRTI, EROTTAA.HAVAITA2, EROTTAA.LUOKITTELU). Skeeman EROTTAA.HAVAITA1 solut eivät ole merkitseviä. Vahvin kytkös elottomiin agentteihin on skeemalla EROTTAA.ERI.

5.3.5 Skeemojen kuvaus analyysin perusteella

Myös *erottaa*-verbin yksimuuttujainen analyysi osoittaa, että verbin ydinargumentteja analysoimalla on mahdollista selvittää merkityksenvaihtelun mekaniikan pääpiirteitä. Analyysin heikkouksia ovat sellaisten skeemojen kuin EROTTAA.ERI ja EROTTAA.HAVAITA1 pieni

kokonaisfrekvenssi, minkä takia tulokset ovat parhaimmillaankin suuntaa antavia. Toinen heikkous on X^2 -testauksen vuoksi tehty skeemojen yhdistelyt, jotka hämärtävät tuloksia jonkin verran. X^2 -testauksen minimiehdot eivät myöskään täyty kaikissa tapauksissa, mikä heikentää tulosten uskottavuutta puhtaasti tilastotieteellisestä näkökulmasta.

Sitä, mitkä muuttujien alaluokista ovat voimakkaimmin läsnä skeemojen välisessä vaihtelussa havainnollistaa taulukko 40. Tähän taulukkoon on seulottu 23:stä yksimuuttujaisessa tarkastelussa mukana olleesta kategoriasta 15 sillä perusteella, kuinka voimakkaasti kolmen eri assosiaation voimakkuusmittarin nojalla i) yksittäiset kategoriat ja skeemat ovat kytköksissä (Cramérin V), ii) skeemat ovat kytköksissä yksittäisiin luokkiin ($U_{\text{muuttujaiskeema}}$) ja iii) yksittäiset kategoriat ovat kytköksissä skeemoihin ($U_{\text{skeema|muuttuja}}$).

Taulukko 40. Yksittäisten luokkien ja skeemojen välisistä assosiaatioista voimakkaimmat.

Cramérin V	$U_{\text{muuttujaiskeema}}$	$U_{\text{skeema muuttuja}}$
X_SEM_ELOTON (0.850) ¹³⁶	X_SEM_ELOTON (0.698)	X_SEM_ELOTON (0.222)
X_SEM_ELOLLINEN (0.850)	X_SEM_ELOLLINEN (0.698)	X_SEM_ELOLLINEN (0.222)
Z_SEM_INSTITUUTIO (0.719)	Z_SEM_INSTITUUTIO (0.485)	Y_SEM_ELOTON (0.178)
Y_SEM_ELOLLINEN (0.708)	Y_SEM_ELOLLINEN (0.424)	Z_SEM_INSTITUUTIO (0.177)
Y_SEM_ELOTON (0.689)	Y_SEM_ELOTON (0.418)	Y_SEM_ELOLLINEN (0.175)
RAKENNE_VAIN_Y (0.650)	RAKENNE_VAIN_Y (0.384)	X_MUOTO_NA (0.145)
X_MUOTO_NA (0.650)	X_MUOTO_NA (0.356)	Z_SEM_ELOTON (0.140)
Z_SEM_ELOTON (0.608)	Z_SEM_ELOTON (0.322)	RAKENNE_VAIN_Y (0.124)
Z_MUOTO_NA (0.472)	Y_SEM_INSTITUUTIO (0.242)	Z_MUOTO_NA (0.068)
Y_SEM_INSTITUUTIO (0.412)	X_MUOTO_NOLLA (0.194)	RAKENNE_KONNEKTORI (0.058)
X_MUOTO_NOLLA (0.407)	RAKENNE_KONNEKTORI (0.188)	X_MUOTO_NOLLA (0.056)
Z_SEM_ELOLLINEN (0.405)	RESIPROOKKI_NA (0.185)	Z_SEM_ELOLLINEN (0.049)
Z_MUOTO_NP (0.391)	RESIPROOKKI_LÄSNÄ (0.185)	Z_MUOTO_NP (0.048)
RAKENNE_KONNEKTORI (0.389)	Z_MUOTO_NA (0.177)	RESIPROOKKI_NA (0.046)
Y_MUOTO_NP (0.344)	Z_SEM_ELOLLINEN (0.172)	RESIPROOKKI_LÄSNÄ (0.046)

Taulukosta havaitaan, että voimakkainta assosiaatio on kaikissa sarakkeissa argumenttien semanttisia ominaisuuksia mittaavissa kategorioissa. Skeemat ennustavat kautta linjan paremmin luokkien käyttäytymistä kuin luokat skeemojen, mikä tarkoittaa, että yksittäisten luokkien tietämisen pohjalta voidaan suhteellisen maltillisesti ennustaa skeemojen taustalla olevaa merkityksenvaihtelua. Symmetrisen Cramérin V :n valossa kärjessä ovat agentin semanttista ontologiaa mittaavat luokat X_SEM_ELOLLINEN~ELOTON ($V=0.850$), jonka jälkeen tulevat Z_SEM_INSTITUUTIO ($V=0.719$), Y_SEM_ELOLLINEN ($V=0.708$), Y_SEM_ELOTON ($V=0.689$) ja lopulta Z_SEM_ELOTON ($V=0.608$). Argumenttirakennetta mittaavista luokista vahvinta assosiaatio on luokilla VAIN_Y ($V=0.650$) ja KONNEKTORI ($V=0.389$), argumenttien läsnäoloa mittaavista kategorioista X_MUOTO_NA ($V=0.650$) ja Z_MUOTO_NA ($V=0.472$).

Yksimuuttujaisen analyysin perusteella ja skeemojen sisäisen jakauman nojalla luokka

¹³⁶ Testaukset ehdot eivät täyty, sillä taulukon soluista 25 %:ssa on arvo > 5 .

X_SEM_ELOTON – joka tarkoittaa, että ilmipantu agentti on eloton – on erityisesti kytköksissä skeemaan EROTTAA.ERI (100 % sisäisesti; *kyky ajatella **erottaa** meidät eläimistä*). Luokkaan X_SEM_ELOLLINEN (agentti on elollinen) taas kytkeytyvät skeemat EROTTAA.LUOKITTELU (91.7 %), EROTTAA.HAVAITA2 (90.8 %; *Psykopaatti on ihminen, jota ei edes "asiantuntijat" pysty **erottamaan** terveiden joukosta*) ja EROTTAA.HAVAITA1 (75 %). Luokka Z_SEM_INSTITUUTIO – joka tarkoittaa, että ilmipannun lähteen spesifioi luokan INSTITUUTIO tarkoite – kytkeytyy skeemoihin EROTTAA.IRT1 (100 %; *Soini lupasi **erottaa** perussuomalaisista kaikki jotka saavat rasistisesta rikoksesta tuomion*) ja EROTTAA.IRT2 (94 %; *Jos kirkko **erotetaan** valtiosta*). Luokkaan Y_SEM_ELOLLINEN – joka tarkoittaa, että patientti on elollinen – kytkeytyvät skeemat EROTTAA.ERI (66.7 %; *Minut **erottaa** perus keskiluokkaisesta suomalaisista se, että iltakaljan sijaan nautin iltajointin*) ja EROTTAA.IRT1 (96.6 %; *Jokaisen puolueen tulee **erottaa** jäsenet, jotka syyllistyvät rasismiin tai muuhun laittomuuteen*). Luokkaan Y_SEM_ELOTON (patientti on eloton) taas kytkeytyvät skeemat EROTTAA.HAVAITA1 (96.4 %; *koira **erottaa** jalkapallo-ottelun ja seuraa palloa*), EROTTAA.HAVAITA2 (79.7 %; *useat eivät **erottaneet** kakkosolutta kolmosesta*) ja EROTTAA.LUOKITTELU (86.7 %; *Paaperon pitäisi nämä käsitteet **erottaa***).

Kategoria RAKENNE_VAIN_Y – joka tarkoittaa argumenttirakennetta (S)VO – kytkeytyy skeemaan EROTTAA.IRT1 (64.7 %; *Eikö Hesari_{subjekti} **voisi** jo **erottaa**_{verbi} tätä surkeaa kolumnistia_{objekti}*). Katteoria RAKENNE_KONNEKTORI (argumenttirakenne (S)VO sisältää kaksi toisiinsa ja-konnektorilla vai vastaavalla rinnastettua jäsentä) kytkeytyy skeemoihin EROTTAA.LUOKITTELU (50.0 %), EROTTAA.IRT2 (42.8 %), EROTTAA.ABSTRAKTI (30.0 %) ja EROTTAA.HAVAITA2 (26.5 %). Agentin poissaoloa tarkasteleva kategoria X_MUOTO_NA kytkeytyy skeemoihin EROTTAA.IRT2 (100 %), EROTTAA.ABSTRAKTI (71.9 %), EROTTAA.KONKRETIA (70.5 %) ja EROTTAA.IRT1 (63.7 %), kun taas nollapersoonaa mittaava kategoria X_MUOTO_NOLLA kytkeytyy skeemoihin EROTTAA.HAVAITA2 (29.0 %), EROTTAA.HAVAITA1 (42.9 %), EROTTAA.LUOKITTELU (41.7 %). Viimein lähteen poissaoloa tarkasteleva kategoria Z_MUOTO_NA skeemoihin EROTTAA.IRT1 (66.1 %) ja EROTTAA.HAVAITA1 (67.9 %).

Jälleen on kiinnostavaa tarkastella automaattisen alamerkityksen tunnistamisen ehtoja eli erottelutarkkuutta p (SKEEMA | LUOKKA) vastakohtana edellisen kappaleen tarkastelulle ja luvun lopun skeemaerottelulle, joiden lähtökohtana on p (LUOKKA | SKEEMA). Tarkkuudella en kuitenkaan tarkoita ennustustodennäköisyyttä (kuten Gries [2006:83–85] ilmeisesti tarkoittaa), vaan prosenttijakaumaa, joka tiettyjen muuttujakriteerien vallitessa vallitsee skeemojen välillä tutkimuskorpuksessa (ks. s. 103). Esitän seuraavaksi joitakin esimerkkejä.

Jos tunnistamisalgoritmi (jolle on syötetty tieto luokkien esiintymisfrekvensseistä skeemojen välillä) havaitsee, että lauseen ilmipannun agentin spesifioi *se*-tukipronomini + lause -rakenne, saavutetaan 100 %:n (n=12) jakauma, että kyseessä on skeema EROTTAA.ERI, ja 81 %:n (n=25) tarkkuus saavutetaan, jos algoritmi havaitsee, että ilmipantu agentti on semanttisesti eloton ja että argumenttirakenteena on SVOA_{ero}. Jos algoritmi havaitsee, että agentti on inkorporoitu verbiin, saavutetaan 65 %:n (n=17) tarkkuus, että kyseessä on skeema EROTTAA.HAVAITA2, ja jos algoritmi havaitsee, että agentti on yksilötarkoitteinen ja että argumenttirakenne on muotoa SVO_{NP+ja+NP} (*Veikko erottaa kissat ja koirat*), saavutetaan 72 %:n jakauma (n=13), joka yhdistyy skeemaan EROTTAA.HAVAITA2. Jos algoritmi tunnistaa, että argumenttirakenne ei sisällä rektioadverbiaalia (Z_MUOTO_NA ~ RAKENNE_VAIN_Y) ja että patientti kuuluu luokkaan YKSILÖ tai JOUKKO ja että X-argumentti eli agentti esiintyy tai ei esiinny ilmipantuna, saavutetaan 92.5 %:n (n=74) jakauma, jonka nojalla kyse on skeemasta EROTTAA.IRTI, ja jos ehdot ovat muuten samat, mutta patientteina saa esiintyä pelkästään YKSILÖ:itä, saavutetaan 96.1 %:n tarkkuus (n=49).

En jatka spekulatiota tämän pitemmälle, vaan esitän lopuksi yksimuuttujaisen analyysin pohjalta kullekin skeemalla kuvauksen.

Skeemoista suurin, EROTTAA.HAVAITA2 (n=155), hakeutuu lauserakenteisiin, joihin on mahdollista koodata kaikki kolme *erottaa*-verbin semanttisen valenssin jäsentä (X-, Y- ja Z-argumentit). Leksikaalisesti ilmipantu tai nollapersoonainen agentti (X_MUOTO_NP + NOLLA) – eli transitiivilauseen tekoketjun ensimmäinen jäsen – on läsnä 92.2 %:ssa (n=134) skeeman korpusesiintymistä. 86.4 %:ssa skeeman korpusesiintymistä täyttyvät monimuuttujaiset ehdot, joissa agentti esiintyy nollapersoonassa tai ilmipantuna ja on semanttiselta ontologialtaan joukko- tai yksilötarkoitteinen (ks. VISK § 1363: ”Nollapersoonaa koskee yksilöä”). Skeema hakeutuukin aktiivilauseisiin (96.1 %, n=149) ja argumenttirakenteisiin i) SVOA_{ero}, ii) SVO_{jaollinen} ja iii) SVO_{NP+ja+NP}, joita X-argumentin (subjektin) osalta vastaa luokka X_SEM_ELOLLINEN ~ X_MUOTO_NOLLA ja Y- ja Z-argumenttien osalta luokat 1) RAKENNE_REKTIO (52.3 %; *Jos poliisi ei erota oikeaa väärästä, niin miksi kansalaisen pitäisi?*), 2) RAKENNE_Y_JAOLLINEN (14.8 %; *Jos yksi rehtori saa keksiä omia rangaistuksiaan [ja] toinen omiaan, on vaikea erottaa, kumpi näistä on nöyryyttävää*) ja 3) RAKENNE_KONNEKTORI (26.5 %; *oppi erottamaan ystävät ja viholliset*). Kategoriaan ELOTON (←KONKRETIA) lukeutuvia tarkoituksia – eli sanoja, jotka eivät poimi henkilöitä ja joukkoja vaan konkreettisia entiteettejä, kuten *tupakka* tai abstrakteja entiteettejä, kuten *lainmukainen toiminta* – on 79.8 %:a Y-argumenttien esiintymistä (y_SEM_ELOTON) ja 81.7 %:a Z-argumenttien esiintymistä (Z_SEM_ELOTON). Ominaisuus luo sukulaisuutta skeemoihin

EROTTAA.LUOKITTELU (Y_SEM_ELTON = 86.6 %, Z_SEM_ELTON = 85.4 %) ja EROTTAA.HAVAITA1 (Y_SEM_ELTON = 96.5 %, Z_SEM_ELTON = 100 %), joista EROTTAA.HAVAITA1 on lähisukulainen Y-argumentin SEM_KONKRETTIA : SEM_ELTON -jakauman kaltaisuuden ja EROTTAA.LUOKITTELU Z-argumentin jakauman kaltaisuuden nojalla.

Skeemassa EROTTAA.HAVAITA2 esiintyy 66.7 %:a (n=22) kaikista *erottaa*-verbin objektin paikalla olevista lausetäydennyksistä (Y_MUOTO_RAKENNE, *erottaa*-verbissä n=33) ja 77.3 % (n=17) objektin rinnasteisen jakson tai rektioadverbiaalin paikalla olevista lausetäydennyksistä tai niiden loppujaksoista (Z_MUOTO_RAKENNE). Molemmat ovat läsnä esimerkiksi *Milläs erotat, mikä rangaistus on nöyryyttävä*_{Y_MUOTO_RAKENNE} *ja mikä* [rangaistus] *ei* [ole nöyryyttävä]_{Z_MUOTO_RAKENNE}. Taipumus lähentää sitä skeemoihin EROTTAA.LUOKITTELU (Z_MUOTO_RAKENNE, n=4) ja EROTTAA.HAVAITA1 (Z_MUOTO_RAKENNE, n=5).

Muuttujan RESIPROOKKI eli *toisistaan*-lekseemin esiintymistä skeemassa on skeemojen välillä eniten (42.9 %, n=30) ja skeemojen sisällä kolmanneksi eniten (19.4 %) skeemojen EROTTAA.LUOKITTELU ja EROTTAA.IRTI2 jälkeen. Tämä ominaisuus (joka yhdessä esiintyy luokkien RAKENNE_KONNEKTORI ~ Y_JAOLLINEN kanssa, ks. s. 118) vihjaa, että skeeman käsitteistyksissä jäsenet ovat siinä määrin tasa-arvoisia, ettei useinkaan – joskaan ei aina¹³⁷ – ole merkitystä sillä, kumpi semanttisen valenssin jäsenistä ensisijaistetaan LÄHDEVÄYLÄ:ä pitkin ”liikkuvaksi” entiteetiksi – edes silloin, kun jäsenet kytketään LÄHDEVÄYLÄ:n tarkentamalla rektioadverbiaalilla, kuten esimerkeissä *Päättäjät kun eivät osaa erottaa sitä, mikä on todella tarpeellista siitä, mikä ei ollenkaan ole* (K362) tai *Jos ostaja ei osaa erottaa vanhaa kalaa tuoreesta, on parasta olla ostamatta kalaa* (K375). Päinvastoin käyttäytyy EROTTAA.IRTI, jossa nimenomaan Y-argumentti spesifioi fokuksena olevan tarkoitteen (’sen joka eroaa’), siis sen, jonka liikkeen tai tilanmuutoksen lause ilmoittaa.

Protolauseen monimuuttujaiset ehdot – jotka voisivat olla: 1) agentin spesifioi ilmipantu yksilö- tai joukkotarkoitteinen NP (X_SEM_ELOLLINEN) tai agentti on nollapersoonassa (X_MUOTO_NOLLA) ja 2) molemmat (sekä Y- että Z-argumentti) ovat mukana merkitysrakenteessa ilmipantuina jollakin kolmesta muuttujan RAKENNE osoittamasta jäsenyysstrategi-

¹³⁷ Elatiivista huolimatta skeeman EROTTAA.HAVAITA2 toteumat ilmoittavat aina staattisen tilan, kuten esimerkki *Psykopaatti on ihminen, jota ei edes "asiantuntijat" pysty erottamaan terveiden joukosta* (K500). Tässä *terveiden joukosta* ilmoittaa viiteryhmän, josta psykopaatit pyritään ikään kuin poimitaan erilleen eli identifioimaan. Lokalistisen hypoteesin (ks. s. 7 alav. 13) mukaan lauseen taustalla voidaan nähdä spatiaaliselta lähdealueelta motivoituvaa metaforista siirtymää, jota kuvastaa lähdealueelle sijoittuva spatiaalinen lause *Erottaa jyvät akanoista ~ akanoiden joukosta*, jossa elatiivi spesifioi LÄHDÄVÄYLÄ:n konkreettisessa merkityksessä. Usein elatiivisijainen rektiotäydennys kuuluu (n=11) tukimuuttujan EROTTAA_Z_TYYPPI (jonka luokkia ovat LÄHDEVÄYLÄ, n=487 ja FOKUS, n=20) luokkaan FOKUS, kuten *Kännikuskien kyllä erottaa ajotavasta* (K197). Tällainen lauseke ilmoittaa ominaisuuden, jonka pohjalta tunnistaminen tapahtuu.

asta (OA_{ero} eli *kissat koirista*, O_{NP+ja+NP} eli *kissat ja koirat* tai O_{jaollinen} *elukat* [kissat ja koirat]) siten, että 3) jäsenten semanttisena luokkana on Y~Z_SEM_ELTON – täyttää 65 %:a (n=100) skeeman korpuresiintymistä. Kiinnostavaa on, että samat ehdot täyttyvät myös 39:ssä eli 65 %:ssa skeeman EROTTAA.LUOKITTELU esiintymistä, 10:ssä (35.7 %) skeeman EROTTAA.HAVAITA1 esiintymistä, kun taas skeemoilla EROTTAA.ERI ~ IRTI ~ IRTI2, n = 0 ja skeemoilla EROTTAA.ABSTRAKTI ~ KONKRETIA, n = 7 (yht.). Skeema toteuttaa Geeraertsin (2006: 141–159) prototyypin kriteereistä paitsi sen, i) että se on taajasti käytössä (skeemoista yleisin, n=155) myös sen, ii) että skeema on funktionaalisesti joustava (spesifikaatioita kaikissa muuttujissa ja luokissa) sekä lisäksi sen, iii) että skeema taipuu erilaisiin merkitystehäviin, joita ovat ainakin 'havaitseminen', 'tunnistaminen', 'entiteettien identifiointi' ja 'tietäminen tai ymmärtäminen' (ks. luku 3). Skeeman semanttisen valenssin jäsenten protoluukumäärä on 3 (X-arg. eli agentti + Y-arg. eli patientti + Z-arg. eli lähde tai pattienkin rinnasteinen jakso), joista kaikki ovat yleensä ilmipantuina (paitsi agentti usein nollana).

Skeema EROTTAA.LUOKITTELU (n=60), joka ilmoittaa älyn tai muun kognitiivisen kyvyn avulla tehtyä entiteettien erottamista, on mitä ilmeisemmin ”sukua” skeemalla EROTTAA.HAVAITA2 – havainto, jonka tein jo skeemojen annotoinnin yhteydessä. Niillä on kuitenkin eroja. Yksi on skeemaa EROTTAA.HAVAITA2 suurempi sisäinen osuus luokassa RAKENNE_KONNEKTORI (50 %, n=30, herkkyydystason 3 merkitsevyys), mikä tarkoittaa, että skeeman toteutumisissa Y- ja Z-argumentit jäsentyvät ensisijaisuuden suhteen neutraaliin muotoon O_{NP+ja+NP} (+ RESIPROOKKI eli *toisistaan*-lekseemi), kuten esimerkissä *Olennaista olisi erottaa toisistaan vapaaehtoisesti harjoitettu prostituutio_{Y-ARG.} ja sitten tämä ihmis-kauppa- ja törkeän parituksen osasto_{Z-ARG.}* (K145). Luokassa RAKENNE_REKTIO (30 %, n=19) LÄHDEVÄYLÄ:n mielikuvaskema (ainakaan spatiaalisessa hengessä) on ehkä EROTTAA.HAVAITA2-skeemakin heikommin läsnä käsitteistyksessä, kuten esimerkissä *Niin, laadustahan tässä puhutaankin, mutta kun sitä ei voi erottaa tuosta pakosta* (K229). Tällä tarkoitan, että entiteetit olisi mahdollista jäsentää Y- ja Z-argumenteiksi kumminpäin tahansa ilman, että käsitteistys isommin muuttuu. Sillä abstraktit oliot, ajatukset tai aatteet, eivät manifestoi ”lokaatiota” spatiaalisten olioiden tavalla (vrt. minimipari *Sipilä pitää erottaa pääministerin pallilta: Pääministerin palli pitää erottaa Sipilästä.*)

Agentin koodauksen suhteen skeemalla EROTTAA.LUOKITTELU on skeemaa EROTTAA.HAVAITA2 isompi sisäinen NA-osuus (X_MUOTO_NA = 18.3 %) ja nollapersoonaosuus (X_MUOTO_NOLLA = 41.7 %), kun taas NP-osuus (40 %) on pienempi. Erot eivät kuitenkaan ole suuria ja voivat mennä ”virhemarginaaliin” piikkiin. Kaikista skeeman korpusesimerkeistä 78 % (n=47) sisältää yksilö- tai joukkotarkoitteisen agentin

(X_MUOTO_NOLLA ~ X_SEM_ELOLLINEN). RESIPROOKKI:a eli *toisistaan*-lekseemiä esiintyy 33.3 %:ssa skeeman korpusesiintymistä, mikä on toiseksi suurin skeemakohtainen osuus skeeman EROTTAA.IRTI2 jälkeen. Elollisen agentin läsnä sekä RESIPROOKKI:n ja toisaalta lausetäydennyksen mahdollisuus (Y_MUOTO_RAKENNE = 8.3 %, n=5; Z_MUOTO_RAKENNE = 9.1 %, n=4) luovat kaltaisuutta skeemaan EROTTAA.HAVAITA2.

Skeema EROTTAA.IRTI2 (n=32), jonka protoesimerkkinä esitettäköön lause *Uskonto ja valtio tulee erottaa toisistaan* (K43), hakeutuu argumenttirakenteisiin VOA_{ero} (RAKENNE_REKTIO = 50 %) ja VONP_{ja+NP} (RAKENNE_KONNEKTORI = 42.8 %) eli käsitteistyksiin, joissa molemmat täydennykset (Y- ja Z-argumentit) on ilmipantuina mutta agentti ilmipanematta (X_MUOTO_NA = 100 %). Y-argumenteista (patientti) 96 %:a ja Z-argumenteista 93.8 %:a kuuluu luokkaan Y~Z_SEM_INSTITUUTIO. Y- ja Z-argumentit spesifioi lähes yksinomaan sellaiset sanat kuin *valtio*, *kirkko* ja *uskonto*. RESIPROOKKI:a esiintyy noin joka toisessa (46.9 %) havaintoesimerkissä. Skeemalla on suurin skeemakohtainen passiivin osuus, 53.1 %:a. Skeema edustaakin agentittomiin rakenteisiin hakeutuvaa ilmaisutyyppiä, jolla puhuja ilmoittaa, että kahden instituution, organisaation tai vastaavan yhteiskunnallispoliittinen kytkös tulisi katkaista. Pikemminkin kuin ”erillistä merkitystä”, ilmaisutyyppi edustaa *erottaa*-verbin eräille skeemoille yleistä skemaattisempaa ”konstruktiota”, jossa kausatiiviketjun ensimmäinen jäsen (agentti eli X-argumentti) on taka-alastettu (taulukossa 38x solulla EROTTAA.MUUT:NA on herkkyystason 3 merkitsevyys). Tässä konstruktiossa ilmipantujen jäsenten protomäärä on 2 (Y-arg. ja Z-arg.): agentti on yleensä taka-alastettu eli läsnä merkitysrakenteessa mutta ilmipanematta lauserakenteessa.

Monimuuttujaiset ehdot, joissa lause on passiivissa ja Y- ja Z-argumentit on ilmipantu (EROTTAA.IRTI2 = 46.9 %), kytkvät sen skeeman sisäisen prosentuaalisen jakauman kaltaisuuden nojalla¹³⁸ sellaisiin skeemoihin kuin EROTTAA.ABSTRAKTI (34.4 %) ja EROTTAA.KONKRETIA (31.8 %) vastakohtana agenttimyönteisille skeemoille EROTTAA.HAVAITA1 (0 %) tai EROTTAA.HAVAITA2 (3.2 %). Kriteerit, joissa pääluokkana on passiivi tai kyseessä on jokin muu agentiton rakenne, kuten nesessiivirakenne, yhdistettynä jommankumman kategorian, RAKENNE_REKTIO:n tai RAKENNE_KONNEKTORI:n, läsnäoloon, täyttää 93.8 % (n=30) skeeman EROTTAA.IRTI2 esiintymistä. Skeemakohtaisen kaltaisuuden nojalla lähimpinä ovat taas EROTTAA.ABSTRAKTI (62.5 %) ja EROTTAA.KONKRETIA (56.8 %) vastakohtana skeemoille EROTTAA.ERI (8.8 %), EROTTAA.HAVAITA1 (0 %) ja EROTTAA.HAVAITA2 (5.8 %).

¹³⁸ Tämä köyhän miehen klusterianalyysi noudattaa mutkat oiottuna klusterianalyysin logiikkaa (ks. luku 6).

Skeemat EROTTAA.ABSTRAKTI (n=32) ja EROTTAA.KONKRETIA (n=44), joista ensimmäisen (proto)toteuma voisi olla lause *Pirkanmaa on aikoinaan erotettu Satakunnasta* (K97) ja jälkimmäisen lause *Tarkoituksena on erottaa virtsa ulosteista* (K177), hakeutuvat lauserakenteisiin, joissa ei esiinny ilmipantua agenttia (kategorian X_MUOTO_NA arvot ovat: EROTTAA.ABSTRAKTI: = 84.4 %, n=27 [passiivia 40 %], EROTTAA.KONKRETIA = 70.5 %, n=31 [passiivia 43 %]). Ominaisuus kytkee ne skeeman EROTTAA.IRTI2 kanssa agenttia karsastavaan konstruktion, jossa ilmipantuna on tyypillisesti vain argumentit Y- ja Z. Skeemassa EROTTAA.KONKRETIA ilmipantuja agenteja on kuitenkin 29.5 %, joista luokkaan ELOLLINEN (JOUKKO ~ YKSILÖ) kuuluvat (54 %) ilmoittavat konkreettista manipulaatiota, kuten lause *Käytännössä poliisi erottaa tappelijat/pahoinpitelijän* (K356).

Skeema EROTTAA.KONKRETIA suosii herkkyystasolla 1 Y- ja Z-argumenttien jäsenyyksessä luokkaa RAKENNE_REKTIO (75 %) – *erottaa virtsa ulosteista* –, tosin sisäisissä prosentiosuuksissa myös skeema EROTTAA.ABSTRAKTI on siihen kallellaan (60 %). Yksimuuttujaisen analyysin tulosten perusteella REKTIO-luokan teoreettisesti odotuksenmukainen jakauma kytkee yhteen skeemoja EROTTAA.ABSTRAKTI, EROTTAA.HAVAITA2 ja EROTTAA.IRTI2, kun taas (herkkyystasolla 1) merkitsevä positiivinen assosiaatio (yliedustus luokassa REKTIO) yhdistää skeemoja EROTTAA.KONKRETIA ja EROTTAA.ERI.

Skeemoja EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA yhdistää lisäksi *toisistaan*-sanan karttaminen (yht. kategoriaa RESIPROOKKI_LÄSNÄ, 6.5 %:a, n=5) ja loitontaa Y- ja Z-argumenttien semanttinen painotus: EROTTAA.KONKRETIA hakeutuu luokkiin Z~Y_SEM_KONKRETIA, kun taas skeema EROTTAA.ABSTRAKTI hakeutuu luokkiin Z~Y_SEM_ABSTRAKTI, tosin myös muut semanttiset luokat tulevat kysymykseen, skeemalla EROTTAA.KONKRETIA erityisesti ihmiset (Y_SEM_ELOLLINEN = 41 %). Skeema EROTTAA.KONKRETIA hahmottaa perfektivisen prosessin, jossa rektioadverbiaali spesifioi LÄHDEVÄYLÄ:n, kuten esimerkiksi *Miten esimerkiksi hampurilaisten jauheliija on erotettu luista* (K178). Lause predikoi jauhelihan (langackerilaisittain [1987] *muuttujan*) siirtymistä lähdeväylää pitkin pois luiden (*kiintopisteen*) yhteydestä. Toisenlaisia ovat skeeman EROTTAA.ABSTRAKTI esimerkit *Vuonna 1919 Lauttasaari erotettiin Helsingin maalaiskunnasta* (K166) tai *vuonna 1992 tuloverolaki, jolla pääomatulojen verotus erotettiin tavallisen kansa ansiotuloista* (K119), jotka ilmoittavat tosimaailmassa tapahtuvaa muutosta – ja hyvin erilaista tilannetta kuin jauhelihaesimerkki¹³⁹. Kategorioiden Z_SEM ja Y_SEM luokkapainotukset ovatkin isoin skeemoja erottava tekijä. Silti en näe syytä, miksi skeemat tulisi katsoa

¹³⁹ Aineistossani on vain yksi skeemaan EROTTAA.ABSTRAKTI koodattu lause, jossa agenttina on elollinen, nimittäin esimerkki *Hän ei ole erottanut Gazaa sen ulkopuolelle* (K342).

”eri merkityksiksi”. Abstrakteimmille dimensioille hahmottuvat tilanteet ovat perustavanlaatuiselta ontologialtaan toisenlaisia kuin spatiaaliseen dimensioon hahmottuvat tilanteet, minkä takia myös niitä predikoivat verbi-ilmaukset poikkeavat toisistaan. Tulkintani mukaan nämä skeemat (ja ehkä lisäksi EROTTAA.IRTI2) kytkeytyvät toisiinsa syntaktisen, argumenttirakenteiden manifestoitumista heijastavan analogian nojalla, joita kutsun skeemakohtaisiksi kolligatiivisiksi profiileiksi (ks. luku 5.5). Isoin ero on siinä, että EROTTAA.ABSTRAKTI ”predikoii” abstrakteja tarkoituksia, EROTTAA.KONKRETIA konkreettisia ja EROTTAA.IRTI2 instituutioita.

Toiseksi suurin skeema EROTTAA.IRTI (n=124; 24.3 % skeemojen välillä), jonka protoesimerkki voisi olla lause *Ihmisten mieliksi joku johtaja voitaisiin varmuuden vuoksi erottaa* (K39), hakeutuu argumenttirakenteeseen (S)VO (RAKENNE_VAIN_Y = 64.7 %; herkkyystason 3 merkitsevyys). Skeeman EROTTAA.HAVAITA1 (RAKENNE_VAIN_Y = 60.7 %) kanssa ne muodostavat Z-argumenttia karsastavan tyypin, jossa argumenttien protolukumäärä on 1–2 (agentti + patientti). Skeeman EROTTAA.IRTI toteumat esiintyvät ilman agenttia 65 %:ssa (n=86) korpusesiintymistä, ja lisäksi sillä on skeemoista toiseksi suurin passiivin osuus skeeman EROTTAA.IRTI2 jälkeen, nimittäin 42 %. Ilmipannun agentin (X_SEM) esiintymät jakautuvat tasan luokkiin X_SEM_ELOLLINEN (n=19; *Soini*) ja X_SEM_INSTITUUTIO (n=19; *eduskunta*)¹⁴⁰. Y-argumentteina (patientteina) esiintyy – ei lausetäydennyksiä tai pronomineja *se* ja *tämä* – vaan lähinnä luokkaan NP (99.2 %) koodattuja tarkoituksia (*talonmies*), jotka ovat semanttisesti (Y_SEM) joukko- (34.7 %) tai yksilötarkoituksisia (61.9 %; elollisia patientteja yhteensä 96.6 %). Lähde – milloin ilmipantu (33.9 %) – lukeutuu aina (100 %) luokkaan INSTITUUTIO (*koulu*, *puolue*). Resiprookkia skeemoissa EROTTAA.IRTI, EROTTAA.HAVAITA1 ja EROTTAA.ERI ei esiinny lainkaan. Vaikka EROTTAA.IRTI esiintyykin mieluummin ilman agenttia – ominaisuus, joka kytkee sen rintamaan EROTTAA.IRTI2 : EROTTAA.ABSTRAKTI : EROTTAA.KONKRETIA – sen voidaan silti nähdä ilmaisevan *erottaa*-verbin ”merkitysverkon ilmaisupotentiaa” yksilöllisesti: herkkyystasolla 3 se karttaa Z-argumentin ilmipanemista ja herkkyystasolla 1 *erottaa*-verbin yleisintä (48.8 %, n=245) Y- ja Z-argumenttien jäsenystapaa OA_{ero} (RAKENNE_REKTIO). Lisäksi se on sillä tavalla semanttisesti ahdasmielinen, että sallii argumenteikseen vain YKSILÖ:itä, JOUKKO:ja, INSTITUUTIO:ita. Ominaisuudet kytkevät sitä skeemaan EROTTAA.IRTI.

erottaa-skeemoista pienin on EROTTAA.HAVAITA1 (n=28), jonka toteuma on lause *Koira erottaa jalkapallo-ottelun ja seuraa palloa* (K276). Sen merkitysrakenteessa esiintyy

¹⁴⁰ Instituutiotarkoitteisen agentin sisältävissä toteutumisissa on lauseita, kuten *Kirkko ei voi erottaa jäseniään seurakunnasta* (K418), joita assosioituvat skeemaan EROTTAA.KIRKKO.

aina agentti joko ilmipantuna (X_MUOTO_NP = 57.1 %) tai nollapersoonassa (X_MUOTO_NOLLA = 42.9 %), mikä samalla tarkoittaa, ettei skeemaa esiinny lainkaan passiivissa. Agenteista 86 % (X_SEM_ELOLLINEN + X_MUOTO_NOLLA) on elollisia (joukko- tai yksilötarkoitteisia). Y-argumentit eli patientit lukeutuvat luokkiin ELOTON (53.6 %) ja KONKRETIA (42.9 %), ilmipannut Z-argumentit (n=9) taas luokkaan KONKRETIA (100 %). Ilmipannuista rektioadverbiaaleista 88 % ei spesifioi ainakaan spatiaalista LÄHDEVÄYLÄ:ä vaan tukimuuttujan EROTTAA_Z_TYYPPI luokan FOKUS ilmoittaman kohteen, kuten lauseessa *Olin aiemmin siinä käsityksessä että eläimet (kuten kissa) eivät erot-taisi mitään kaksiulotteisista kuvista tai esim. telkkarista*_{Z_ARG} (K277). Tässä Z-argumentti spesifioi visuaalisen ankkuripisteen, jonne havaitsemisen stimulus ikään kuin lokalisoituu. Lisäksi skeemassa esiintyy lausetäydennyksiä (18 %) niin kuin esimerkissä *Todennäköisesti ne erottavat, mitä kuvissa on, koska niiden katse pysyy pitempään koirien kuin vaikkapa lelujen päällä* (K275). Kaikki edellä mainitut ominaisuudet kytkevät sitä skeemaan EROTTAA.HAVAITA2. Skeemasta EROTTAA.HAVAITA2 loitommalta ja kohti skeemaa EROTTAA.IRT1 sitä kuitenkin kytkee valenssiltaan kaksipaikkaisen argumenttirakenteen, SVO, suosiminen (X_MUOTO_NP = 57.1 %; RAKENNE_VAIN_Y = 60.7 %). Skeema EROTTAA.HAVAITA1 poikkeaa muista eroskeemoista (ei kuitenkaan täysin skeemasta EROTTAA.HAVAITA2) siinä, että se voi esiintyä myös *semanttiselta* valenssiltaan kaksipaikkaisena (rakenteessa ”X tekee aistihavainnon Y”) – toisin kuin skeema EROTTAA.IRT1, jonka merkitystaustaan kuuluu lähes aina implisiittinen lähde. Skeemasta EROTTAA.IRT1 poiketen sen merkitysrakenteessa on lisäksi aina henkilötarkoitteinen agentti, mikä tarkoittaa että semanttisen valenssin jäsenten protolukumäärä on 2 (agentti + patientti). Analyysin pohjalta on vaikea sanoa, minkä skeeman yhteyteen se luontevammin sijoittuu, ehkä skeeman EROTTAA.HAVAITA2. Koodausprosessissa skeeman EROTTAA.HAVAITA1 esimerkeiksi on päätynyt myös lauseita, jotka ilmoittavat pikemminkin mentaalista toimintaa (sopien paremmin skeemaan EROTTAA.HAVAITA2) kuin puhdasta havaitsemista. Niiden esimerkkien osalta, jotka ilmoittavat ”puhdasta” havaintoa *nähdä-* tai *maistaa-*verbien tavoin, skeeman voi ajatella edustavan itsenäistä merkitystä, ’sitä että agentti tekee aistihavainnon’. Tällainen merkityskomponentti on yleensä vaikuttamassa skeeman EROTTAA.HAVAITA2 havainnoissa, tosin ei ”puhtaana” vaan erilaisten mentaalisten prosessien – kuten tunnistamisen – taustalla.

Viimeisen skeeman, EROTTAA.ERI (n=34), jota esiintyy vain aktiivilauseissa, protoesi-merkki voisi olla *Minut **erottaa** perus keskiluokkaisesta suomalaisista lähinnä se, että ilta-kaljan sijaan nautin iltajointin* (K262). Skeemojen sisällä sillä on suurin luokan

RAKENNE_REKTIO eli jäsenyyksen VOA_{ero} jakauma (85.3 %), joka on merkitsevä herkkyystasolla 1. Agenttia esiintyy 91.2 %:ssa esimerkeistä, ja kaikki niistä edustavat luokkaa ELOTON. Erityisen usein skeemassa esiintyy luokkien PRON (n=7) ja RAKENNE (n=12; 38 %:a ilmipannuista agenteista) tarkoituksia eli *se*, *tämä* ja *joka*-edussanaisia lausekkeita ja (tuki-pronomillisia⁰) lausetäydennyksiä. Luokan X_MUOTO_PRON (←RAKENNE) kytkös onkin merkitsevä herkkyystasolla 3. Tulokset eivät ole yllättäviä, sillä skeeman koodauskriteerinä on toiminut, että merkitysrakenteessa esiintyy elementti, joka sopii sanakirjojen antamaan alamerkitysparafraasiin 'tehdä erilaiseksi, muista poikkeavaksi'. Agenttitarkoite onkin 93.3 %:ssa luokan VASTAUS spesifikaatio. Y- ja Z-argumenttien spesifikaatiot hakeutuvat luokkiin Z~Y_SEM_ELOLLINEN. Etenkin Z-argumentin kytkös yksilö- ja joukkotarkoituksiin (yht. 59.4 %) on merkitsevä herkkyystasolla 3 luoden poikkeavaa profiilia suhteessa muihin skeemoihin.

On hankala sanoa, minkä *erottaa*-skeeman sukulainen EROTTAA.ERI on, jos minkään. Ainakin ominaisuus suosia ilmipantuja elottomia agenteja erottaa sen muista *erottaa*-verbin skeemoista¹⁴¹. Kategorian RAKENNE_REKTIO suosiminen kytkee sitä skeemaan EROTTAA.KONKRETIA vastakohtana rektioadverbiaalikkammoisille skeemoille EROTTAA.IRT1 ja EROTTAA.HAVAITA1. Toisaalta joukkotarkoitteisten patienttien suosiminen kytkee sitä juuri skeemaan EROTTAA.IRT1. En pohdi asiaa sen enempää vaan siirryn koonnin kautta *eronominin* yksimuuttujaiseen analyysiin.

5.3.6 Yhteenveto ja monimuuttujainen näkökulma

erottaa-skeemojen voidaan ryhmitellä monella tapaa, esimerkiksi paikkaluvun nojalla. Ensimmäinen ryhmä, johon kuuluu EROTTAA.IRT1, on tyypillisesti 1-paikkainen (patientti) tai 2-paikkainen (agentti + patientti). Se hakeutuu argumenttirakenteeseen (S)VO, jossa verbin ydinargumenteista ilmipantuna on yleensä vain patientti. Transitivisen tekoketjun ensimmäinen jäsen, kausatiivinen manipuloija, on kuitenkin mukana käsitteistyksessä taka-alais-tettuna. Toiseen tyyppiin kuuluu EROTTAA.HAVAITA1, joka on tyypillisesti 2-paikkainen (agentti + patientti). Se hakeutuu argumenttirakenteeseen SVO, jossa agentin spesifioi ilmipantu elollinen tarkoite tai agentti on nollapersoonassa¹⁴². Kolmanteen tyyppiin lukeutuvat passiivimyönteiset skeemat EROTTAA.ABSTRAKTI, EROTTAA.KONKRETIA ja EROTTAA.IRT2,

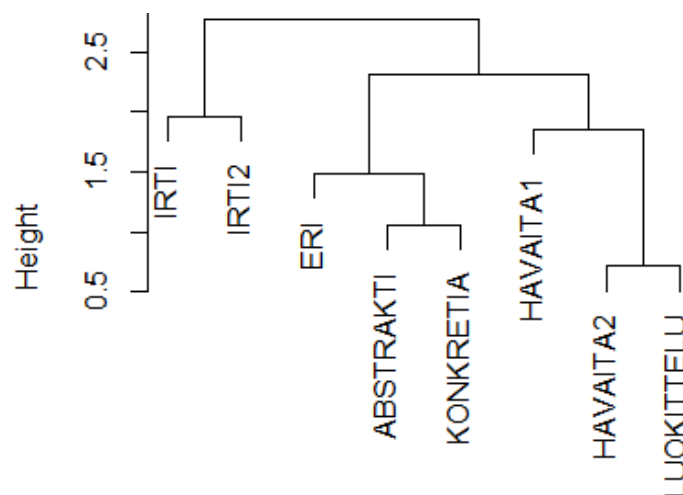
¹⁴¹ Voidaan myös pohtia, hahmottavatko mikroskeemaan EROTTAA.VÄLILLÄ (n=3) toteumat ehkä samankaltaisen, staattisen tilan kuitenkin sillä erotuksella, että kyse on spatiaaliseen, ei abstraktiin tilaan sijoittuvasta käsitteistyksestä, kuten kvasiresultatiivisessa lauseessa *Radioaktiivista uraania sisältävän malmin erottaa pohjavesistä alle ranteen paksuinen kumimatto*_{X_ARG}. Ehkä hahmottavat.

¹⁴² S signaloi tässä myös ilmipanematonta agenttia, joka on aktiivisesti läsnä käsitteistyksessä.

jotka esiintyvät yleensä 2-paikkaisina (patientti + lähde tai patientin rinnasteinen jakso) ja hakeutuvat argumenttirakenteisiin VOA_{ero} , $VO_{NP+ja+NP}$ sekä $VO_{jaollinen}$. Kolmanteen tyyppiin kuuluvat skeemat $EROTTA.AHAVAITA2$ ja $EROTTA.LUOKITTELU$, jotka suosivat 3-paikkaista jäsennystä (agentti + patientti + lähde tai patientin rinnasteinen jakso). Näiden agentteina esiintyy elollisia tarkoituksia. Neljäs tyyppi, $EROTTA.ERI$, esiintyy sekin 3-paikkaisena, mutta agentin spesifioi tyypillisesti eloton olento.

Agenttien valossa *erottaa*-aineistosta hahmottuu neljä roolia: eloton agentti ($EROTTA.ERI$, $EROTTA.ABSTRAKTI$), aistiva- tai ajatteleva agentti ($EROTTA.AHAVAITA1$, $EROTTA.AHAVAITA2$, $EROTTA.LUOKITTELU$), fyysisesti manipuloiva agentti ($EROTTA.KONKRETIA$) sekä sosiaalisesti manipuloiva agentti ($EROTTA.ERTI$).

Yksimuuttujaisen analyysin pohjalta on hankala tehdä tätä pitemmälle meneviä johtopäätöksiä skeemojen yhteyksistä, ja nämäkin ovat spekulatiivisia. Selvää on, argumenttien ja skeemojen välillä on kytköksellisyttä. Kuten Arppe (mts. 73) huomauttaa, tällaiset yksittäiset piirteet puskevat esiin usein niin ilmiselvästi, että saattaa olla houkutus yksinkertaistaa koko tutkittava ilmiö niiden varaan. Ideaalitapauksessa yksimuuttujaisen analyysin tueksi saadaan monimuuttujaisista tukea. Esitänkin lopuksi klusterianalyysikuvan. Klusteroitavana ovat *erottaa*-skeemat, joiden käyttäytymisprofileihin on syötetty yksimuuttujaisessa analyysissä mukana olleet luokat samoine luokkayhdistelyine kuin X^2 -testeissä¹⁴³.



Kuvio 6. *erottaa*-verbin klusterianalyysi. 39 ID-merkkiä. Etäisyys: *euclidean*, Metodi: *Ward*.

Kuviosta 6, jota luetaan alhaalta ylöspäin (ks. luku 6), vihjaa, että skeemat $EROTTA.AHAVAITA2$ ja $EROTTA.LUOKITTELU$, jotka yhdistyvät ensimmäisenä, ovat skeemoista samankaltaisimpia. Ylempänä niihin yhdistyy skeema $EROTTA.AHAVAITA1$, mikä heijastanee sitä, että sen ja skeemojen $EROTTA.AHAVAITA2$ ja $EROTTA.LUOKITTELU$ välillä on

¹⁴³ Luokkayhdistelyt ovat lepusumat kuin X^2 -testeissä, ja mukana on myös pienifrekvenssisia luokkia.

samankaltaisuutta, mutta myös eroavaisuuksia. Keskelle muodostuu klusteri EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA, joka sekin tukee yksimuuttujaisen analyysin tuloksia. Ylemmällä tasolla siihen yhdistyy skeema EROTTAA.ERI, mikä tarkoittaa, että samankaltaisuus on skemaattisempaa. Näitä kolmea skeemaa yhdistää analyysin perusteella etenkin se, että niiden toteumissa on ilmipantuina lähes aina sekä objekti että rektioadverbiaali, eli Y- ja Z-argumentit. Lisäksi analyysi niputtaa yhteen skeemat EROTTAA.IRTI ja EROTTAA.IRTI2. Ne kuitenkin yhdistyvät viimeisenä, mikä tarkoittaa, että niiden välinen samankaltaisuus on vähäisempää kuin vaikkapa skeemojen EROTTAA.HAVAITA2 ja EROTTAA.LUOKITTELU. Tulkinta on, että klusteristruktuuri vahvistaa jo tehtyjä päätelmiä. Isoin ero on skeeman EROTTAA.IRTI2 yhteys skeemaan EROTTAA.IRTI. Kytös ei kuitenkaan ole yllätys, sillä skeema EROTTAA.IRTI2 on koodausvaiheessa lohkaistu juuri skeemasta EROTTAA.IRTI2 (ks. luku 3.2.1). En jatka analyysiä tämän pidemmälle, vaan siirryn näistä alustavista tuloksista *ero*-nominiin ja palaan *erottaa*-verbiin yhteenvedossa 5.5

5.4 *ero*-nomini

Ennen varsinaisen analyysin aloittamista on todeta *ero*-aineiston ”sekakoosteisuus”. Noministatuksensa takia *ero*-lekseemin sisältäviä lausekkeita (n=594) esiintyy aineistossani:

- i) transitiivilauseen subjektina (n=5), objektina (163) ja lauseobjektina (37),
- ii) intransitiivilauseen subjektina (n=51) sekä
- iii) kopulalauseen subjektina tai predikatiivina (n=338).

Lisäksi on pidettävä mielessä, että *ero*-aineisto on koostettu kahdesta sijamuodosta, jotka ovat nominatiivi (ilmiasu = *ero*, n=365) ja partitiivi (ilmiasu = *eroa*, n=229). Ne esiintyvät 8 skeemassa, jotka ovat ERO.ERI, ERO.IRTI, ERO.PUOLISO, ERO.NUMEROT, ERO.HAVAITA, ERO.YMMÄRTÄÄ, ERO.TEHDÄ.ERO ja ERO.NÄHDÄ.

Luodakseni kuvaa aineistosta tarkastelen taulukossa 41 skeemojen jakaumaa suhteessa lauseenjäsenäisyysroolia mittaavaan muuttujaan LAUSE, jonka alakategorioita toimivat OBJEKTI, LAUSEOBJEKTI ja SUBJ./PRED. (sillä en erotellut subjektin ja predikatiivin lauseenjäsenäisyysasemaa). Yhdistin taulukkoa varten luokat ERO.TEHDÄ.ERO ja ERO.NÄHDÄ (näin olen toiminut myös muissa *ero*-analyysissä). Lihavoituna ovat skeemakohtaisesti suurimmat arvot. Taulukossa 41b on esitetty *ero*- ja *eroa*-ilmiasujen jakauma (muuttuja ERO_SIJA), taulukossa 41c muuttujan LAUSETYYPPI jakauma, joka spesifioi informaation siitä, onko kyseessä intransitiivi-, transitiivi- vai kopulalause.

Taulukko 41. *ero*-skeemat ja niiden jakauma lauseenjäsennystehtävittäin (muuttuja LAUSE).

	ERI	NUMEROT	PUOLISO	IRTI	TEHDÄ	YMMÄRT.	HAVI.	Σ_f
SUBJ./PRED.	311	43	27	10	1	2	0	394
OBJEKTI	13	2	9	26	52	36	25	163
LAUSEOBJEKTI	9	2	1	0	1	23	1	37
Σ_f	333	47	37	36	54	61	26	594

Taulukko 41b. Muuttujan ERO_SIJA jakauma skeemojen kesken.

	ERI	NUMEROT	PUOLISO	IRTI	TEHDÄ	YMMÄRT.	HAVI.	Σ_f
<i>ero</i>	248	43	30	11	12	17	4	365
<i>eroa</i>	85	4	7	25	42	44	22	229

Taulukko 41c. Muuttujan LAUSETYYPII jakauma.

	ERI	NUMEROT	PUOLISO	IRTI	TEHDÄ	YMMÄRT.	HAVI.	Σ_f
INTRANSITIIVI	21	12	11	4	1	2	0	51
KOPULA	290	30	14	4	0	0	0	338
TRANSITIIVI	22	5	12	28	53	59	26	205

Taulukoista havaitaan, että etenkin skeemat ERO.ERI ja ERO.NUMEROT suosivat nominatiivia (*ero*) ja hakeutuvat kopulalauseisiin. Päinvastoin käyttäytyvät skeemat ERO.IRTI, ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ ja ERO.HAVAITA, jotka hakeutuvat transitiivilauseisiin ja etenkin kolme viimeistä – ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ ja ERO.HAVAITA – nimenomaan objektin lauseenjäsennyspaikalle. Tästä johtuen ne suosivat partitiivia (*eroa*). Skeema ERO.PUOLISO esiintyy tasaisesti kaikissa kolmessa lausetyypissä.

Tämän prologin jälkeen aloitan varsinaisen analyysin *ero*-nominin ensimmäistä täydennystä mittaavista muuttujista Y_MUOTO ja Y_SEM.

5.4.1 Muuttujat Y_MUOTO ja Y_SEM

Testeissä muuttujan Y_MUOTO alakategoriat on yhdistetty siten, että mukana ovat NA (n=159), NP (\leftarrow PERS, n=4; yht. n=328), PRON (\leftarrow PRON, n=62; RAKENNE, n=42). Jakauma on esillä taulukossa 42. Näin yhdistellyt kategoriat mittaavat oikeastaan vain sitä, esiintyykö rakenteessa ilmipantu Y-argumentti vai ei ja jos esiintyy, kuuluuko se luokkaan NP vai jompaankumpaan luokista PRON ja RAKENNE.

Taulukko 42. Muuttujan Y_MUOTO esiintymät suhteessa *ero*-lekseihin skeemoihin.

	ERI	TEHDÄ	YMMÄRT.	IRTI	HAVI.	PUOLISO	NUMEROT	Σ_m
NA	77	7	8	7	10	28	22	159
NP	185	36	44	28	10	9	20	332
PRON	71	11	9	1	6	0	5	103
Σ_s	333	54	61	36	26	37	47	594

Havaitaan, että vasemmanpuoleisissa skeemoissa yleisin luokka on NP, kun taas skeemoissa ERO.PUOLISO ja ERO.NUMEROT yleisin on NA. Skeema ERO.HAVAITA painottuu yhtä paljon molempiin.

Koska testisuuri, X^2 ($df=12$, $p = 5.427e^{-13}$) = 84.602, ylittää khiin neliön kriittisen arvon X^2 ($df=12$, $\alpha=0.05$) = 21.026, ja koska p ($p = 5.427e^{-13}$) alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujien välillä on kytkös, joka on voimakkuudeltaan kohtalaista ($V=0.267$). Assosiaatio on vähän vahvempaa skeemoista luokkia kohtaan ($U_{\text{muuttuja|skeema}}=0.07$, $\lambda_{\text{muuttuja|skeema}}=0.08$) kuin luokista skeemoja kohtaan ($U_{\text{skeema|muuttuja}}=0.04$, $\lambda_{\text{skeema|muuttuja}}=0$), joskin kumpaankin suuntaan heikohkoa. Poikkeavuuksista tarkempaa tietoa saadaan taulukon 42c avulla. Se esittää havaittujen ja odotettujen arvojen erotuksen, merkitsevyyden ja suunnan (määritettynä taulukon 42x tietojen pohjalta, ks. liite 1).

Taulukko 42c. Muuttujan Y_MUOTO erotustaulukko. Herkkyystaso 1 (vaalealla taustalla) ja herkkyystaso 3 (tummalla taustalla). Merkitsevät havainnot on lihavoitu.

	ERI	TEHDÄ	YMMÄRT.	HAV1.	IRTI	PUOLISO	NUMEROT
NA	-/12.1	-/7.5	-/8.3	+/3.0	-/2.6	+/18.1	+/9.4
NP	-/1.1	+/5.8	+/9.9	-/4.5	+/7.9	-/11.7	-/6.3
PRON	+/13.3	+/1.6	-/1.6	+/1.5	-/5.2	-/6.4	-/3.1

Taulukkoja 42c ja 42x silmäilemällä havaitaan yksi herkkyystason 3 löytö, joka on skeeman ERO.PUOLISO positiivinen yhteys ($X^2 = 33.1$, ks. liite 1) luokkaan NA (+18.1). Herkkyystasolla 1 se karsastaa luokkia NP ja PRON. Herkkyystasolla 1 myös skeema ERO.PUOLISO suosii luokkaa NA (+9.4 esiintymää). Tulkinta on, että ERO.NUMEROT ja etenkin ERO.PUOLISO suosivat lauserakenteita, joissa ei esiinny ilmipantua ensimmäistä täydennystä. Luokkaan NP positiivisesti kytköksissä ovat skeemat ERO.IRTI (+7.9 esiintymää) ja ERO.YMMÄRTÄÄ (+9.9 esiintymää), joista jälkimmäinen skeeman ERO.TEHDÄ.ERO kanssa karsastaa luokkaa NA.

Lisätietoa Y-argumentin suhteesta skeemoihin voidaan yrittää saada tarkastelemalla muuttujaa Y_SEM. Testauksen takia poistin luokan NA ($n=158$) ja yhdistin muuttujan kategorioiksi ELTON (\leftarrow KONKRETIA, $n=46$) ja ELOLLINEN (\leftarrow INSTITUUTIO, $n=28$; JOUKKO $n=46$ ja YKSIÖ $n=57$). Taulukko 43 esittää muuttujan Y_SEM jakauman.

Taulukko 43. Muuttujan Y_SEM jakauma ($n=436$).

	ERI	TEHDÄ	YMMÄRT.	HAV1.	IRTI	PUOLISO	NUMEROT	\sum_m
ELTON	191	37	41	15	0	0	21	305
ELOLLINEN	56	10	12	1	29	10	4	131
\sum_s	256	47	53	16	29	10	25	436

Taulukosta 43 havaitaan, että luokka ELTON on valta-asemassa skeemoissa ERO.ERI, ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ, ERO.HAVAITA sekä ERO.NUMEROT. Päinvastoin käyttäytyvät skeemat ERO.IRTI ja ERO.PUOLISO, joissa vain luokalla ELOLLINEN on esiintymiä. Niiden ensimmäisenä, intransitiivilauseen agenttia vastaavana täydennyksenä toimii siis luokkiin YKSIÖ, JOUKKO ja INSTITUUTIO koodattuja tarkoituksia.

Koska¹⁴⁴ testisuure, $X^2 (df=6, p < 2.2e^{-16}) = 103.2$, ylittää khiin neliön kriittinen arvon $X^2 (df=6, \alpha=0.05) = 12.592$, ja koska $p (p < 2.2e^{-16})$ alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujien välillä on kytkös, joka on melko voimakas ($V=0.506$). Assosiaatio on selvästi vahvempaa skeemoista luokkia kohtaan ($U_{\text{muuttuja|skeema}}=0.218$, $\lambda_{\text{muuttuja|skeema}}=0.319$) kuin luokista skeemoja kohtaan ($U_{\text{skeema|muuttuja}}=0.094$, $\lambda_{\text{skeema|muuttuja}}=0$). Tämä tarkoittaa, että luokkien pohjalta voidaan heikosti ennustaa, mistä skeemasta on kyse. Taulukko 43c esittää havaittujen ja odotettujen arvojen erotuksen, merkitsevyyden ja suunnan herkkyytasoilla 1 ja 3 (merkitsevyydet määrittyvät taulukosta 43x, ks. liite 1).

Taulukko 43c. Muuttujan Y_SEM odotettujen ja havaittujen arvojen erotus, suunta ja merkitsevyys herkkyytasoilla 3 (tummalla taustalla) ja 1 (vaalealla taustalla). Tilastollisesti merkitsevät solut on lihavoitu.

	ERI	TEHDÄ	YMMÄRTÄÄ	HAV1.	IRTI	PUOLISO	NUMEROT
ELOTON	+11.9	+4.1	+3.9	+3.8	-20.3	-7	+3.5
ELOLLINEN	-11.9	-4.1	-3.9	-3.8	+20.3	+7	-3.5

Taulukosta 43c havaitaan kolme herkkyytason 3 havaintoa. Niitä ovat skeeman ERO.IRTI negatiivinen kytkös luokkaan ELOTON ($X^2=-20.3$, ks. liite 1) ja erityisesti positiivinen kytkös luokkaan ELOLLINEN ($X^2=+47.2$, ks. liite 1), johon myös skeema ERO.PUOLISO on kytköksissä ($X^2=+16.3$, ks. liite 1). Herkkyytason 1 tuloksia silmäillessä paljastuu kaksi ryhmää, joista ensimmäisen muodostavat edelliset eli skeemat ERO.IRTI ja ERO.PUOLISO, jotka karsastavat elottomia Y-argumentteja. Toiseen kuuluvat skeemat ERO.ERI ja ERO.HAVAITA jotka ovat positiivisesti kytköksissä luokkaan ELOTON ja negatiivisesti luokkaan ELOLLINEN. Vaikka loppujen skeemojen tulokset eivät ole merkitseviä edes herkkyytasoilla 1, skeemojen voidaan varovaisesti arvioiden katsoa kuuluvan jälkimmäiseen ryhmään, joka suosii Y-argumentteja, jotka on koodattu elottomiksi.

Muuttujilla LUOKKA ja TARKKA ei ole khiin neliö -testien valossa merkitsevää yhteyttä skeemoihin. Jatkan tarkastelemalla muuttujia Z_MUOTO ja Z_SEM, jotka mittaavat *ero*-nomin toista täydennystä tai seuralaista.

5.4.2 Muuttujat Z_MUOTO ja Z_SEM

Testausta varten muuttujan Z_MUOTO alakategorioiksi on yhdistetty luokiksi NP (\leftarrow PRON2, $n=2$; RAKENNE, $n=38$) ja NA ($n=270$). Täten muuttuja mittaa sitä, esiintyykö Z-argumentti lauserakenteessa vai ei. Taulukko 44 esittää muuttujan jakauman ($n=594$).

¹⁴⁴ Soluja > 5 on yhteensä 2 eli 14 % yhteensä 14:sta solusta, jolloin X^2 -testin ehdot täyttyvät.

Taulukko 44. Muuttujan Z_MUOTO jakauma (n=594).

	ERI	TEHDÄ	YMMÄRT.	HAV1.	IRTI	PUOLISO	NUMEROT	\sum_m
NP	203	44	47	9	14	0	7	324
NA	130	10	14	17	22	37	40	270
\sum_s	333	54	61	26	36	37	47	594

Taulukkoa 44 silmäilemällä havaitaan selvästi kaksi ryhmää. Ensimmäinen ryhmä, johon kuuluvat skeemat ERO.ERI, ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ, suosii ilmikoodattuja Z argumentteja. Jälkimmäinen, johon kuuluvat skeemat ERO.HAVAITA, ERO.IRTI, ERO.PUOLISO ja ERO.NUMEROT puolestaan suosii ilmikoodaamattomia Z-argumentteja. Koska testisuure, X^2 ($df=6$, $p < 2.2e^{-16}$) = 115.720, ylittää khiin neliön kriittinen arvon X^2 ($df=6$, $\alpha=0.05$) = 12.592, ja koska p ($p < 2.2e^{-16}$) alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujien välillä on kytkös, joka on lähes melko voimakasta ($V=0.441$). Assosiaatio on selvästi vahvempaa skeemoista luokkia kohtaan ($U_{\text{muuttuja|skeema}}=0.164$, $\lambda_{\text{muuttuja|skeema}}=0.320$) kuin luokista skeemoja kohtaan ($U_{\text{skeema|muuttuja}}=0.077$, $\lambda_{\text{skeema|muuttuja}}=0$). Tuloksista voidaan tehdä se johtopäätös, että luokan tietäminen antaa vain vähän eväitä ennustaa skeema oikein. Poikkeamien suuntaa, merkitsevyyttä ja sijaintia tarkastelee taulukko 44c, jonka soluinformaatio määrittäyty liitteen 1 taulukon 44x avulla.

Taulukko 44c. Muuttujan Z_MUOTO odotettujen ja havaittujen arvojen erotus, suunta ja merkitsevyys herkkyystasoilla 3 (tumma tausta) ja 1 (vaalea tausta). Merkitsevät havainnot on lihavoitu.

	ERI	TEHDÄ	YMMÄRTÄÄ	HAV1.	IRTI	PUOLISO	NUMEROT
NP	+21.4	+14.5	+13.7	-5.2	-5.6	-20.2	-18.6
NA	-21.4	-14.5	-13.7	+5.2	+5.6	+20.2	+18.6

Taulukon 44c tulokset vahvistavat jo tehtyjä havaintoja. Erityisen merkitseviä eli herkkyystason 3 havaintoja on neljä kappaletta ja ne kaikki sijoittuvat skeemoihin ERO.PUOLISO ja ERO.NUMEROT. Niillä on negatiivinen kytkös luokkaan NP ja positiivinen luokkaan NA, joista jälkimmäinen on X^2 -jakauman pohjalta painavampi (ERO.PUOLISO, $X^2=24.2$; ERO.NUMEROT, $X^2=16.3$, ks. liite 1). Skeeman ERO.IRTI havainnot eivät ole tilastollisesti merkitseviä.

Lisätietoa voidaan saada tutkimalla muuttujaa Z_SEM, jonka jakauma näkyy taulukossa 45. Luokiksi muodostuivat NA-luokan (n=260) poistamisen jälkeen ELOTON (\leftarrow KONKRETIA, n=40) ja ELOLLINEN (\leftarrow INSTITUUTIO, n=35; JOUKKO, n=21 ja YKSILO, n=34).

Taulukko 45. Muuttujan Z_SEM jakauma (n=324).

	ERI	TEHDÄ	YMMÄRT.	HAV1.	IRTI	PUOLISO	NUMEROT	\sum_f
ELOTON	150	36	35	8	0	0	5	234
ELOLLINEN	53	8	12	1	14	0	2	90
\sum_f	203	44	47	9	14	0	7	324

Taulukkoa 45 silmäilemällä havaitaan oikeastaan kolme ryhmää. Ensimmäinen, johon kuuluu tällä kertaa myös skeema ERO.HAVAITA sekä ERO.NUMEROT, suosii ilmikoodatuissa Z-argumenteissa luokkaa ELOTON. Skeema ERO.IRTI käyttäytyy päinvastoin: siinä esiintyy ainoastaan kategoriaan ELOLLINEN koodattuja Z-argumentteja. Skeemalla ERO.PUOLISO ei esiinny ilmikoodattuja Z-argumentteja. Yhdistettynä Y-argumentin tuloksiin voidaan todeta, että sille tyypillistä on esiintyä ilman täydennyksiä. Taulukon X^2 -testaus ei ole mahdollista kuin rajuin luokkayhdistelyin, minkä takia se on jätetty tekemättä. Ainoa kiinnostava löytö on skeeman ERO.IRTI ja luokan ELOLLINEN yhteys, mikä onkin täysin odotuksenmukainen.

5.4.3 Muuttujat X_SEM ja X_MUOTO

Muuttujat X_MUOTO ja X_SEM mittaavat sellaisten transitiivilauseiden agenttia, joissa *ero*-lauseke esiintyy objektina (n=200); intransitiivi- ja kopulalauseissa muuttujat koodautuvat luokkaan NA:ksi. Taulukko 46 esittää muuttujan jakauman luokkajaolla NP (\leftarrow PRON), PERS (\leftarrow INKORPOROITU), NOLLA ja NA.

Taulukko 46. Muuttujan X_MUOTO jakauma (n=594).

	ERI	TEHDÄ	YMMÄRT.	HAVL.	IRTI	PUOLISO	NUMEROT	Σ_f
NP	4	15	27	5	0	0	0	49
PERS	2	18	16	7	0	0	0	44
NOLLA	1	9	4	9	0	0	0	23
NA	327	12	14	5	36	37	47	478
Σ_f	333	54	61	26	36	37	47	594

Taulukkoa 46 silmäilemällä havaitaan selviä tendenssejä. Silmiinpistävintä on, että skeemoissa ERO.IRTI, ERO.PUOLISO ja ERO.NUMEROT (sekä käytännössä myös skeemassa ERO.ERI) ei esiinny X-argumentteja, mistä voidaan päätellä, etteivät skeemat esiinny transitiivilauseissa¹⁴⁵. Jäljelle jäävien skeemojen erot ovat todella pieniä: yhdistelmä skeema ERO.TEHDÄ.ERO suosii inkorporoituja ja persoonapronominilla spesifioituja agentteja, skeema ERO.YMMÄRTÄÄ NP-muotoisia ja skeema ERO.HAVAITA nollapersoonaisia agentteja.

X^2 -testausta varten kategoriat on järjestettävä dikotomisesti luokkiin NP (\leftarrow PERS) ja NA (\leftarrow NOLLA). Taulukon X^2 -testaus ei tosin tuo paljoa lisätietoa siihen, mitä jo tiedämme skeemojen jakautumisesta lausetyyppien välille. Koska testisuure, X^2 ($df=6$, $p < 2.2e^{-16}$) = 314.503, ylittää khiin neljän kriittinen arvon X^2 ($df=6$, $\alpha=0.05$) = 12.592, ja koska p ($p < 2.2e^{-16}$) alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujien välillä on

¹⁴⁵ Luokka X_MUOTO_NA ei tee eroa transitiivilauseen koodaamattoman, implisiittisen agenttikonstituentin (esim. passiivi) ja intransitiivi- tai kopulalauseen (koskaan) esiintymättömän agenttikonstituentin välillä. Tätä tarkoitusta varten klusterianalyysissä on luokka NA.TRANS, joka poimii transitiivilauseiden NA-agentit erilleen.

assosiaatiota, joka on melko voimakasta ($V=0.727$). Voidaan myös testata, onko transitiivilauseissa esiintyvien skeemojen – ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ ja ERO.HAVAITA – välillä eroja. Koska näiden ja muuttujan X_MUOTO ristiintaulukoinnin X^2 -testisuure, X^2 ($df=2$, $p = 0.679$) = 0.07734, alittaa khiin neliön kriittisen arvon X^2 ($df=2$, $\alpha=0.05$) = 5.99, ja koska p ($p = 0.679$) ylittää riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujan välillä ei ole tilastollisesti merkitsevää riippuvuutta.

Tarkastelen lopuksi taulukkoa 47, joka esittää muuttujan X_SEM jakauman, josta on poistettu NA ($n=478$) ja yhdistetty loput luokat kategorioihin ELOLLINEN ja ELOTON.

Taulukko 47. Muuttujan X_SEM jakauma ($n=116$).

	ERI ₂	TEHDÄ	YMMÄRT.	HAV1.	IRTI	PUOLISO	NUMEROT	\sum_f
ELOTON	4	1	0	0	0	0	0	3
ELOLLINEN	2	41	47	21	0	0	0	113
\sum_f	6	42	47	21	0	0	0	116

Silmäilemällä taulukko 47 havaitaan, että muuttujan X_SEM kategoria ELOLLINEN esiintymät painottuvat skeemoihin, joissa ylipäänsä esiintyy transitiivilauseen agenttia. Kategorialla ELOTON on esiintymiä 3. Tilastollinen testaus on mahdoton, mutta ilman sitäkin voidaan arvata, ettei muuttuja X_SEM ole merkitsevässä roolissa tutkimassani ilmiössä.

5.4.4 Muuttujat RAKENNE ja RESIPROOKKI

ero-lekseemin kohdalla mittasin argumenttirakennetta kahdella muuttujalla, jotka ovat RAKENNE ja ERO_SIIA. Muuttujan RAKENNE saa kuusi ala-arvoa, jotka ovat REKTIO ($n=14$ ← VAIN_Z, $n=6$), REKTIO2 ($n=56$), VAIN_Y ($n=31$), Y_JAOLLINEN ($n=114$), KONNEKTORI ($n=255$) ja NA ($n=124$). Taulukossa 48 on esillä jakauma ja 48b:ssä prosentuaaliset jakaumat.

Taulukko 48. Muuttujan RAKENNE jakauma ($n=594$).

	ERI	TEHDÄ	YMMÄRT.	HAV1.	IRTI	PUOLISO	NUMEROT	\sum_f
KONNEKTORI	159	40	45	6	1	0	4	255
Y_JAOLLINEN	74	6	7	8	1	2	16	114
NA	55	4	7	7	1	28	21	124
VAIN_Y	1	0	0	2	19	7	3	31
REKTIO2	44	4	2	3	0	0	3	56
REKTIO	0	0	0	0	14	0	0	14
\sum_f	333	54	61	26	36	37	47	594

Taulukosta 48 paljastuu joitakin kiinnostavia havaintoja. Ensinnäkin luokalla REKTIO ($n=14$) – joka tarkoittaa, että argumenttirakenteessa esiintyy Y-argumentti ja elatiivisijaista Z-argumenttia – on esiintymiä ainoastaan skeemassa ERO.IRTI, johon keskittyvät myös luokan VAIN_Y esiintymät. Tämä havainto on siksi kiinnostava, että nämä *ero*-lekseemissä harvinaisimmat luokat – REKTIO ja VAIN_Y – ovat *erota*- ja *erottaa*-verbissä yleisimpien luokkien

joukossa (ks. s. 114). Lisäksi luokka VAIN_Y kytkeytyy nimenomaan skeemoihin EROTA.IRTI ja EROTTAA.IRTI. *ero*-nominin yleisimmät luokat – NA (n=124) ja KONNEKTORI (n=255) – puolestaan ovat *ero*-verbeissä kaikista harvinaisimpia. *ero*-nominissa KONNEKTORI:a suosivat erityisesti skeemat ERO.ERI, ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ, luokkaa NA taas skeemat ERO.PUOLISO ja ERO.NUMEROT. Skeema ERO.HAVAITA osoittaa jälleen hajontaa eri luokkien kesken muodostaen jo eriteltyjen kolmen ryhmittymän lisäksi ikään kuin neljännen ryhmän, jossa esiintymähuiput jakautuvat luokkiin KONNEKTORI (n=6), NA (n=7) ja Y_JAOLLINEN (n=8). Skeemalla ERO.HAVAITA on kuitenkin niin vähän esiintymiä, ettei sen pohjalta voida tehdä kuin hyvin alustavia johtopäätelmiä. Luokka REKTIO2 on ominaisin skeemalle ERO.ERI.

X^2 -testausta varten soluja on liian paljon. Siksi joko skeemoja tai luokkia on yhdisteltävä. Yhdistämismahdollisuuksia on monia, joista taulukon 48 perusteella vähiten jakaumaa vääristävältä tuntuu luokkien KONNEKTORI, Y_JAOLLINEN ja REKTIO2 yhdistäminen luokkaan MUUT. Testauksen minimiehtojen vuoksi on lisäksi yhdistettävä luokat VAIN_Y ja REKTIO luokaksi REKTIO_VAIN_Y. Tämäkään ei vielä riitä, vaan myös skeemoja on karsittava esimerkiksi siten, että intransitiiviseen *erota*-verbiin assosioituvat skeemat IRTI ja PUOLISO yhdistetään kategoriaan (IRTI_PUOLISO) ja transitiiviseen *erottaa*-verbiin assosioituvat ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ yhdistetään (TEHDÄ.YMMÄRTÄÄ). Nyt testaus voi alkaa.

Koska testisuure, X^2 ($df=8$, $p < 2.2e^{-16}$) = 332.285, ylittää khiin neliön kriittisen arvon X^2 ($df=8$, $\alpha=0.05$) = 15.507, ja koska p ($p < 2.2e^{-16}$) alittaa riskitason ($\alpha=0.05$), voidaan päätellä, että skeemojen ja muuttujien välillä on assosiaatiota, joka on melko voimakasta ($V=0.529$). Assosiaatio on jonkin verran vahvempaa skeemoista luokkia kohden ($U_{\text{muuttuja|skeema}}=0.297$, $\lambda_{\text{muuttuja|skeema}}=0.207$) kuin luokista skeemoja kohden ($U_{\text{skeema|muuttuja}}=0.183$, $\lambda_{\text{skeema|muuttuja}}=0.146$). Tämä tarkoittaa, että tietämällä skeeman voidaan jälleen paremmin (eli reilussa yhdessä tapauksessa viidestä) ennustaa, mikä luokka tulee kysymykseen kuin päinvastoin. Poikkeamien suuntaa, merkitsevyyttä ja sijaintia tarkastelee taulukko 48c, jonka so-luinformaatio määrittyy liitteen 1 taulukon 48x avulla.

Taulukko 48c. Muuttujan RAKENNE odotettujen ja havaittujen arvojen erotus, suunta ja merkitsevyys herkkyytasoilla 3 (tummalla taustalla) ja 1 (vaalealla taustalla). Merkitsevät solut on lihavoitu.

	ERI	TEHDÄ_YMMÄRTÄÄ	HAVAITA	IRTI_PUOLISO	NUMEROT
REKTIO_VAIN_Y	-/24.2	-/8.7	+/0.0	+/33.5	-/0.6
NA	-/14.5	-/13.0	+/1.6	+/14.8	+/11.2
MUUT	+/38.7	+/21.7	-/1.6	-/48.2	-/10.6

Taulukosta 48c havaitaan kolme herkkyytason 3 havaintoa, joista painavimmat ovat yhdistelmäskeeman IRTI_PUOLISO positiivinen kytkös ($X^2=+202.6$, ks. liite 1) luokkaan

REKTIO_VAIN_Y (+33.5 esiintymää) sekä negatiivinen kytkös ($X^2=-44.5$, ks. liite 1) luokkaan MUUT (-48.2 esiintymää). Viimeinen herkkyytason 3 havainto on edellisen vastapariksi asettava skeeman ERO.ERI negatiivinen yhteys ($X^2=-23.3$, ks. liite 1) luokkaan REKTIO_VAIN_Y (-24.2 esiintymää). Herkkyytastasolla 1 skeemat ERO.ERI sekä ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ muodostavat yhtenäiseltä vaikuttavan ryhmän. Ne suosivat argumenttirakenteita, joissa sekä Y- että Z-argumentti ovat ilmipantuina. Jälleen teoreettista odotusjakaumaa parhaiten mukailee ERO.HAVAITA, jolla ei ole yhtään merkitsevää solua. Skeema ERO.NUMEROT kytkeytyy luokan MUUT karsastamisen (-10.6 esiintymää) ja luokan NA suosimisen (+11.2 esiintymää) kautta skeemoihin ERO.IRTI ja ERO.PUOLISO. Tulokinta on, että nimenomaan nämä kolme skeemaa esiintyvät argumenttirakenteissa, jossa ei ole ilmipantuina kumpaakaan *ero*-nominin täydennyksistä. Erityisesti ERO.PUOLISO vaikuttaa sillä tavalla konseptuaalisesti autonomiselta (ks. Leino 1993: 83), että sen ilmoittama entiteettien välinen relaatio (*ero ~ eroaminen*, 'parisuhteen päättyminen') voidaan käsitteistää, ilman että käsitteistetään relaation kytkemät entiteetit ('*he, joita ero koskee*'), joskin ne voidaan myös ilmipanna (Y_JAOLLINEN, n=2; VAIN_Y, n=7), jolloin ne ainakin ovat käsitteistyksessä (ks. luku 5.4.5).

Toinen *ero*-nominin argumenttirakenteen jäsentymistä mittaava muuttuja on ERO_SIIA_KAIKKI, josta on puhetta sivulla 46. Alakategorioiden huomattavan määränsä vuoksi sen tilastollinen analyysi ei ole relevanttia. Tarkastelenkin tässä muuttujaa ERO_SIIA johon yhdistin luokkia havainnoimalla niiden jakauman suhteessa *ero*-skeemoihin. Luokka MUUT (n=161) sisältää luokat NOMINATIIVI, LAUSE, PARTITIIVI, NA. Luokka GENETIIVI_ELATIIVI (n=52) luokat GENETIIVI sekä luokat joiden nimessä esiintyy ELATIIVI. Kolmas luokka OLOSIJA (n=381) sisältää jäljelle jääneet luokat, joista isoimmat ovat GENETIIVI_ADESSIIVI, ADESSIIVI ja INESSIIVI. Taulukossa 49 on esillä muuttujan SIIA_KOOSTE ja taulukossa 49b prosentuaalinen jakauma.

Taulukko 49. Muuttujan ERO_SIIA jakauma (n=594).

	ERI	TEHDÄ	YMMÄRT.	HAVI.	IRTI	PUOLISO	NUMEROT	\sum_f
ELATIIVI	6	1	3	1	30	4	7	52
MUUT	74	5	10	12	5	32	23	161
OLOSIJA	253	48	48	13	1	1	18	381
\sum_f	333	54	61	26	36	37	47	594

Taulukko 49b. Muuttujan ERO_SIIA prosentuaalinen jakauma skeemojen välillä (vaalea tausta) ja skeemojen sisällä (tumma tausta).

	ERI		TEHDÄ		YMMÄRT.		HAVI.		IRTI		PUOLISO		NUMEROT	
ELATIIVI	11.5	1.8	1.9	1.9	5.8	4.9	1.9	3.8	57.7	83.3	7.7	10.8	13.5	14.9
MUUT	46.0	22.2	3.1	9.3	6.2	16.4	7.5	46.2	3.1	13.9	19.9	86.5	14.3	48.9
OLOSIJA	66.4	76.0	12.6	88.9	12.6	78.7	3.4	50.0	0.3	2.8	0.3	2.7	4.5	36.2

Silmäilemällä taulukkoja havaitaan tendenssejä. Skeeman ERO.PUOLISO esiintymät painottuvat luokkaan MUUT (85.5 %), kun taas skeeman ERO.IRTI esiintymät painottuvat luokkaan ELATIIVI (83.3 %). Nämä skeemat myös karttavat luokkaa OLOSIA, jota suosivat erityisesti skeemat ERO.ERI (76.0 %), ERO.TEHDÄ.ERO (88.9 %) ja ERO.YMMÄRTÄÄ (78.7 %). Skeemojen ERO.HAVAITA ja ERO.NUMEROT esiintymät jakautuvat erityisesti luokkien MUU ja OLOSIA kesken. Tulokset vahvistavat jo tehtyjä havaintoja. Kun niitä ja muuttujan SIIA tietoja analysoidaan, selviää, että skeeman ERO.IRTI ilmipantu Y-argumentti esiintyy aina joko genetiivissä (76.8 %) tai nominatiivissa (23.2 %) ja Z-argumentti joko elatiivissa (93.3 %) ja genetiivissä (6.7 %). Skeeman ERO.PUOLISO Y-argumentti taas esiintyy vain genetiivissä (50 %), nominatiivissa (37.5 %) tai adessiivissä (12.5 %).

Ennen yhteenvedon tarkastelen vielä lopuksi muuttujaa RESIPROOKKI (n=120), jonka jakauma on esillä taulukossa 50. Taulukossa 50c on esillä muuttujien RAKENNE ja RESIPROOKKI ristiintaulukointi.

Taulukko 50c. Muuttujan RESIPROOKKI ja RAKENNE ristiintaulukointi.

	VAIN_Z	NA	VAIN_Y	REKTIO	REKTIO2	Y_JAOLLINEN	KONNEKTORI
NA	6	124	31	8	47	93	164
RESIPROOKKI	0	0	0	0	9	21	91

Taulukko 50. Muuttujan RESIPROOKKI jakauma skeemoittain (n=594).

	ERI	TEHDÄ	YMMÄRT.	HAVI.	IRTI	PUOLISO	NUMEROT	\sum_m
NA	265	23	47	22	36	37	43	473
RESIPROOKKI	68	31	14	4	0	0	4	121
\sum_s	333	54	61	26	36	37	47	594

Havaitaan, että *ero*-nominin yhteydessä muuttujan RAKENNE yhdistelmäluokan MUUT kanssa eli luokkien REKTIO2 (n=9), Y_JAOLLINEN (n=21) ja KONNEKTORI (n=89), missä se muistuttaa sekä *erota*- että *erottaa*-verbien käyttäytymistä. Lisäksi havaitaan, että muuttujalla RESIPROOKKI ei ole lainkaan esiintymiä skeemoissa ERO.PUOLISO ja ERO.IRTI, missä se muistuttaa *erota*-verbin skeemoja EROTA.PUOLISO ja EROTA.IRTI sekä *erottaa*-verbin skeemaa EROTTAA.IRTI. Resiprookkia esiintyy jonkin verran skeemoissa ERO.YMMÄRTÄÄ (23 %) ja ERO.ERI (20.4 %) sekä lisäksi skeemoissa ERO.HAVAITA (15.4 %) ja ERO.NUMEROT (8.5 %). Ainoastaan skeema ERO.TEHDÄ.ERO esiintyy useammin (57.4 %) muuttujan RESIPROOKKI kanssa kuin ilman. Tässä se assosioituu sellaisiin resiprookki-myönteisiin skeemoihin EROTTAA.LUOKITTELU ja EROTTAA.HAVAITA2.

5.4.5 Yhteenvedo

Keskeisin havainto *ero*-nominin yksimuuttujaisesta analyysistä on, etteivät *erottaa*- ja *erota*-verbin esitutkimuksen pohjalta syntyneet, X-, Y- ja Z-argumenttien ominaisuuksia mittaavat

muuttujat selittäneet tutkittavaa ilmiötä yhtä hyvin kuin *ero*-verbien tapauksessa. Silti analyysi paljastaa, että *ero*-skeemojen välillä on eroja, joita pystytään paikantamaan jopa hyvin karkeilla ELOTON : ELOLLINEN tai rakenteessa : ei-rakenteessa -jaotteluilla. X^2 -testauksen rajoitusten vuoksi tehdyt luokkajärjestelyt saattavat jonkin verran hämärtää tuloksia.

Taulukkoon 51 on koostettu 17 analyysissä mukana olleesta luokasta sillä perusteella, kuinka voimakkaasti kolmen eri assosiaation voimakkuusmittarin nojalla i) yksittäiset luokat ja skeemat ovat kytköksissä (Cramérin V), ii) skeemat ovat kytköksissä yksittäisiin luokkiin ($U_{\text{muuttuja|skeema}}$) ja iii) yksittäiset luokat ovat kytköksissä skeemoihin ($U_{\text{skeema|muuttuja}}$). Mukana ovat johdannon muuttujat LUOKKA, LAUSEFUNKTIO ja ERO_SIJA. Taulukosta on tummennettuna luokat, jotka osoittavat jossakussa sarakkeessa voimakkainta assosiaatiota.

Taulukko 51. Yksittäisten luokkien ja skeemojen välinen assosiaation voimakkuus tarkasteltuna kolmella mittarilla.

Cramérin V	$U_{\text{skeema muuttuja}}$	$U_{\text{muuttuja skeema}}$
LAUSE_ERO_OBJEKTII (0.828)	LUOKKA_TRANSITIIVI (0.265)	RAKENNE_REKTIO (0.652)
LAUSE_ERO_SUBPRED (0.828)	LAUSE_ERO_OBJEKTII (0.264)	LAUSE_ERO_OBJEKTII (0.601)
LUOKKA_TRANSITIIVI (0.828)	LAUSE_ERO_SUBPRED (0.264)	LAUSE_ERO_SUBPRED (0.601)
RAKENNE_REKTIO (0.801) ¹⁴⁶	LUOKKA_KOPULA (0.243)	LUOKKA_TRANSITIIVI (0.6)
LUOKKA_KOPULA (0.764)	X_MUOTO_NP (0.163)	X_MUOTO_NP (0.546)
X_MUOTO_NP (0.728)	X_MUOTO_NA (0.163)	X_MUOTO_NA (0.546)
X_MUOTO_NA (0.728)	RAKENNE_REKTIO (0.12)	LUOKKA_KOPULA (0.517)
SIJA_GENETIIVI_ELAT (0.685) ¹⁴⁷	RAKENNE_MUUT (0.119)	SIJA_GENETIIVI_ELAT (0.433)
RAKENNE_MUUT (0.597)	SIJA_OLOSIJA_ILLATIIVI (0.111)	RAKENNE_MUUT (0.29)
SIJA_OLOSIJA_ILLATIIVI (0.548)	Y_SEM_ELOTON (0.091)	SIJA_OLOSIJA_ILLATIIVI (0.247)
ERO_SIJA_ERO (0.502)	Y_SEM_ELOLLINEN (0.091)	Y_SEM_ELOTON (0.203)
ERO_SIJA_EROA (0.502)	ERO_SIJA_ERO (0.089)	Y_SEM_ELOLLINEN (0.203)
Y_SEM_ELOTON (0.487)	ERO_SIJA_EROA (0.089)	ERO_SIJA_ERO (0.195)
Y_SEM_ELOLLINEN (0.487)	RAKENNE_KONNEKTORI (0.089)	ERO_SIJA_EROA (0.195)
RAKENNE_KONNEKTORI (0.462)	SIJA_GENETIIVI_ELATIIVI (0.088)	RAKENNE_KONNEKTORI (0.19)
Z_MUOTO_NA (0.441)	Z_MUOTO_NA (0.078)	Z_MUOTO_NA (0.164)
Z_MUOTO_NP (0.441)	Z_MUOTO_NP (0.078)	Z_MUOTO_NP (0.164)

Koostetta silmäilemästä havaitaan, että kaikkien mittarien valossa voimakkaimpien kategorioiden joukossa ovat LAUSE_ERO_OBJEKTII~ERO_SUBPRED ($V = 0.828$), jotka tarkastelevat, onko *ero*-lauseke lauseenjäsenyksellisesti objekti vai subjekti tai predikaatti sekä kategoriat LUOKKA_TRANSITIIVI ($V = 0.828$) ja LUOKKA_KOPULA ($V = 0.764$), jotka mittaavat oikeastaan samaa asiaa, nimittäin sitä, kuuluuko *ero*-lauseke transitiivilauseeseen vai kopulalauseeseen vaikutusalaan. Tietämällä, mikä skeema on kysymyksessä, pystytään kolmessa tapauksessa viidestä ennustamaan, onko kyse transitiivilauseesta vai ei (LUOKKA_TRANSITIIVI, $U_{\text{muut-}}$

¹⁴⁶ X^2 -testissä soluista 42 % on alle 5. Yksittäisten luokkien testausta varten suoritin muuttujalle RAKENNE kaksi testiä: yhden samoilla luokilla kuin luvussa 5.4.4. ja toisen yhdistämällä vain luokat REKTIO \leftarrow VAIN_Z.

¹⁴⁷ Khiin neliö -testissä soluista 35.7 % on alle 5.

tuja|skeema=0.6), ja jos tiedetään, että kyse on transitiivilauseesta, voidaan noin yhdessä tapauksessa viidestä kertoa, mistä skeemasta on kyse ($U_{\text{skeema|muuttuja}}=0.265$). Argumenttien läsnäoloa mittaavista luokista kärjessä ovat X_MUOTO_NP ja X_MUOTO_NA ($V=0.728$), jotka mittaavat paitsi transitiivilauseen agentin läsnäoloa, samalla dikotomiaa transitiivilause : kopula tai intransitiivilause. Y- ja Z- argumenttien jäsennystä tarkastelevista luokista kärjessä ovat luokat RAKENNE_REKTIO ($V=0.801$), jota vastaa lause *Useat kunnat suunnittelevat eroa kuntaliitosta* (K1210), RAKENNE_MUUT ($V=0.597$) – joka sisältää luokat KONNEKTORI, Y_REKTIO2 (*Siinä siis se ratkaiseva ero satukirjoihin*, K1338) ja Y_JAOLLINEN (*Viineissä on myös eroa*, K1623) – sekä RAKENNE_KONNEKTORI ($V=0.462$), jota vastaa lause *On tärkeää tehda ero vapaaehtoisen ja pakotetun seksityön välille* (K1615). Semanttista ontologiaa mittaavista luokista kärjessä ovat kategoriat Y_SEM_ELTON ~ ELOLLINEN (0.487), jotka tarkastelevat, onko *ero*-lekseemin ensimmäinen täydennys (Y-argumentti) elollinen vai eloton.

Jälleen sarake 2 ($U_{\text{skeema|muuttuja}}$), joka ilmoittaa kytköstä yksittäistä luokista skeemoihin päin, on muihin sarakkeisiin verrattuna heikointa kautta linjan. *ero*-verbien analyysissä mukana olleista yksittäisistä kategorioista vahvin kytkös on transitiivisen agentin läsnäoloa mittaavan muuttujan X_MUOTO kategoriat NP ja NA ($U_{\text{skeema|muuttuja}}=0.163$), jota seuraavat muuttujan RAKENNE luokat REKTIO ja MUUT. Muuttujan Y_SEM luokkien ELTON ~ ELOLLINEN assosiaatio on heikkoa ($U_{\text{skeema|muuttuja}}=0.091$), samoin muuttujan Z_MUOTO luokkien NP ja NA ($U_{\text{skeema|muuttuja}}=0.078$). Vaikka mittaustulokset ovat heikohkoja, täytyy muistaa, että todellisessa kielenkäytössä yksittäiset kielenpiirteet (joita muuttujien alakategoriat mittaavat, olkoonkin että karkeasti) kytkeytyvät toisiin lukuisin tavoin, tai kuten Arppe (2008: 256) asian ilmoittaa: “features are multiply intercorrelated in real linguistics usage”. Yksimuuttujaisen analyysin heikkous onkin juuri se, ettei siinä pystytä ottamaan huomioon eri muuttujien yhteisvaikutusta.

Jälleen on kiinnostavaa tarkastella automaattisen alamerkityksen tunnistamisen ehtoja suunnassa p (SKEEMA | LUOKKA). Esitän joitakin esimerkkejä. Jos tunnistamisalgoritmi, johon on syötetty tieto skeemojen esiintymistaajuudesta ja luokkien esiintymistaajuudesta skeemojen välillä, yksinkertaisesti tulkitsee jokaisen *ero*-nominin esiintymän skeeman ERO.ERI toteumaksi, saavutetaan aineistossani 56 % tarkkuus, sillä 594 *ero*-esimerkistä 333 kuuluu skeemaan ERO.ERI. Jos *ero* esiintyy nominatiivissa (ERO_SIJA_ERO), saavutetaan 67.9 %:n tarkkuus. Koska skeeman esiintymät kattavat yksin yli puolet aineistostani, luokatkin painottuvat skeemojen välillä siihen. Suurimmat skeemojen väliset osuudet ovat luokilla RAKENNE_REKTIO2 (78.6 %), Z_SEM_JOUKKO (76.2 %), Z_MUOTO_RAKENNE (73.7 %; Väitän, että *on ero sillä, ampuuko jonkun hengiltä vai käveleekö päin punaisia suoja tiellä*,

K1554), SIJA_OLOSIJA (66 %, *Puuttumisella ja huomioimisella on vissi ero*, K1526), RAKENNE_KONNEKTORI ~ Y_JAOLLINEN (63 %) ja Y_SEM_ELOTON ~ KONKRETIA (62.6 %).

Vielä korkeampiin tunnistustarkkuuksiin päästäisiin, jos algoritmi havaitsisi, millaisesta lausetyypistä on kyse. Tunnistaessa, että kyseessä on kopulalause (LUOKKA_KOPULA, johon skeemojen välillä lukeutuu 85 % skeeman ERO.ERI havainnoista) ja lisäksi että muuttujan rakenne luokkana on joko REKTIO2, Y_JAOLLINEN tai KONNEKTORI, saavutetaan 93.7 % (n=240) tarkkuus, ja jos algoritmi havaitsee, että kyse on kopulalauseesta ja että *ero*-nominin toinen täydennys (Z_SEM) on luokaltaan ELOTON tai KONKRETIA, saavutetaan 98.5 % (n=140) tarkkuus. Jos algoritmi sen sijaan tunnistaa, että *ero*-nominin ilmipantuna on vain Y-argumentti (RAKENNE_VAIN_Y), jonka spesifioi elollinen olio, saavutetaan 83 % tarkkuus, että kyseessä on skeema ERO.IRTI, ja jos systeemi havaitsee, että *ero*-nominin Z-argumenttina esiintyy erosijainen täydennys (RAKENNE_REKTIO~VAIN_Z), saavutetaan 100 %:n (n=14) tarkkuus. 100 %:n jakaumiin päästään siten, että algoritmi havaitsee, että *ero*-nomini esiintyy *tehdä*-, *ymmärtää*- tai *huomata*-verbien objektina, sillä skeemat ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ ja ERO.HAVAITA on määritelty perusteella verbi + objekti (*ero*). Erotteleviin jakaumiin päästäisiin myös tarkastelemalla (yhdessä joidenkin muiden muuttujien kanssa) muuttujan ERO_SIJA_KAIKKI alakategorioiden ja skeemojen suhdetta – sitä, minkälaisissa sijamuotojen kombinaatioissa *ero*-nominin seuralaiset esiintymät. Lopetan spekulaa-tion tähän ja annan *ero*-skeemojen protoluonnehdinnat määrittelysuunnassa *p* (LUOKKA | SKEEMA).

Skeema ERO.ERI – jonka protototeuma voisi olla lause *Koiralla ja rotalla nyt sattuu olemaan sellainen melko ratkaiseva ero sisuskaluissa, että [...]* (K1448) – on esiintymätajuutensa (n=329) nojalla *ero*-nominin prototyyppi. Sen esiintymät painottuvat kopulalauseisiin (87 %) ja nominatiiviin eli *ero*-muotoon (74 %). Se hakeutuu argumenttirakenteisiin, joita edustavat muuttujan RAKENNE luokat KONNEKTORI (47.7 %), Y_JAOLLINEN (22.2 %) ja REKTIO2 (13.2 %). Herkkyystasolla 1 ERO.ERI muodostaakin skeemojen ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ kanssa rintaman, jotka suosivat argumenttirakenteita, joissa sekä Y- että Z-argumentti on ilmipantuina (ERO.ERI, Y_MUOTO_LÄSNÄ¹⁴⁸ = 76.9 %; Z_MUOTO_LÄSNÄ = 60.7 %). Oppositioksi niille muodostuu Z-argumenttia karsastava joukko, johon luokan Z_MUOTO_NA taipumuksen nojalla kuuluvat ERO.PUOLISO (100 %, herkkyystaso 3), ERO.NUMEROT (85.1 %, herkkyystaso 3), ERO.HAVAITA (65.4 %) ja ERO.IRTI (61.1 %, herkkyystaso 1). Semanttisesti skeeman Y- ja Z-argumentit painottuvat luokkiin ELOTON ~ KONKRETIA (Y_SEM = 74.6 %; Z_SEM = 73.9 %). Ominaisuus kytkee sen Y-argumentteina luokkien ELOTON ja KONKRETIA tarkoitteita suosivaan piiriin, johon kuuluvat skeemat

¹⁴⁸ Muuttujan Y_MUOTO kategorioihin pohjautuva tukimuuttuja, joka kertoo, onko argumentti läsnä.

ERO.HAVAITA (93.7 %, herkkyytaso 3), ERO.NUMEROT (84 %), ERO.TEHDÄ.ERO (78.7 %) ja ERO.YMMÄRTÄÄ (77.4 %). Lause täydennyksiä skeemassa esiintyy erityisesti Z-argumenteina (Z_MUOTO_RAKENNE, skeemojen välillä 73.7 % ja skeeman sisäisesti 8.4 %, n=28). Resiprookin (esim. välillä) esiintymiä skeemalla ERO.ERI on skeemojen välillä eniten (56.2 %) ja skeemojen sisällä kolmanneksi eniten (20.4 %, n=68). Monet näistä ominaispiirteistä muistuttavat skeeman EROTA.ERI ominaispiirteitä – asia, johon palaan luvuissa 5 ja 6.

Skeema ERO.NUMEROT (n=46) – jonka toteuma on lause *10 pisteen ero on oikeustieteellisen pääsykokeessa valtava* (K1384) – yhdessä esiintyy luokkien RAKENNE_NA (44.7 %) ja RAKENNE_Y_JAOLLINEN (34 %) kanssa eli ilman täydennyksiä tai jaollisen täydennyksen kera. Sen esiintymät hakeutuvat (yht. 89.4 %:ssa) kopula- (n=30) ja intransitiivilauseisiin (n=12). Ilmipantuja X-argumentteja (transitiivilauseen agentteja) ei esiinny ollenkaan. Semanttisesti ilmipannut Y- (n=25; 53.2 %) ja Z-argumentit (n=7; 14.9 %) edustavat luokkaa eloton (Y_SEM_ELOTON 84 %; Z_SEM_ELOTON 71.4 %). Ne ovat siis numeroita. Skeema suosii *ero*-muotoa (91.5 %, n=43), samoin kuin skeema ERO.PUOLISO (81.1 %, n=30).

Myös skeema ERO.PUOLISO (n=37) – jonka malliesimerkki on *Ehkä ero onkin positiivinen asia, joka helpottaa kaikkien elämää* (K1662) – suosii herkkyytetasolla 1 luokka RAKENNE_NA (75.7 %), joka tarkoittaa, ettei kumpaakaan Y- ja Z-argumenteista ole ilmipantu. Z-argumentteja ei esiinny aineistossani kertaakaan. Ilmipannut Y-argumentit (n=10) ovat herkkyytetasolla 3 joko joukko- (n=7) tai yksilötarkoitteisia (n=3). Tulokset ovat samankaltaisia kuin skeemalla EROTA.PUOLISO, joka suosii argumenttirakennetta V (RAKENNE_NA), ja Z-argumentin eli lähteen poisjättöä sekä joukkotarkoitteisia agentteja. Resiprookkia ei esiinny skeemoissa EROTA.PUOLISO ja ERO.PUOLISO, missä ne muistuttavat skeemoja EROTTAA.ERI, EROTA.IRTI ja ERO.IRTI, jossa ei myöskään esiinny RESIPROOKKI:a.

Skeema ERO.IRTI (n=36) – jonka malliesimerkki on *Lähes kaksi promillea aikuisväestöstä vaatii ministerin eroa* (K1209) – hakeutuu ainoana skeemana *erota-* ja *erottaa-*verbeihin assosioituviiin luokkiin RAKENNE_REKTIO (skeemojen välillä 100 % ja sisällä 39.9 %) ja RAKENNE_VAIN_Y (skeemojen välillä 58.1 % ja skeeman sisällä 50 %). Ilmipannut Z-argumentit (n=14) ovat aina luokkaan INSTITUUTIO (n=12) koodattuja tarkoituksia tai yksiköitä ja joukkoja (n=2). Y-argumentit ovat herkkyytetasolla 3 elollisia, tyypillisesti yksilöitä (n=18; 62.8 %), mikä erottaa sen skeemasta ERO.PUOLISO, joka suosii joukkotarkoitteisia Y-argumentteja. Skeeman toteumissa on kyse yksilötarkoitteisten olentojen eroamisesta, missä se assosioituu skeemoihin EROTA ~ EROTTAA.IRTI, joiden Y-argumentiksi hakeutuu yksilöitä.

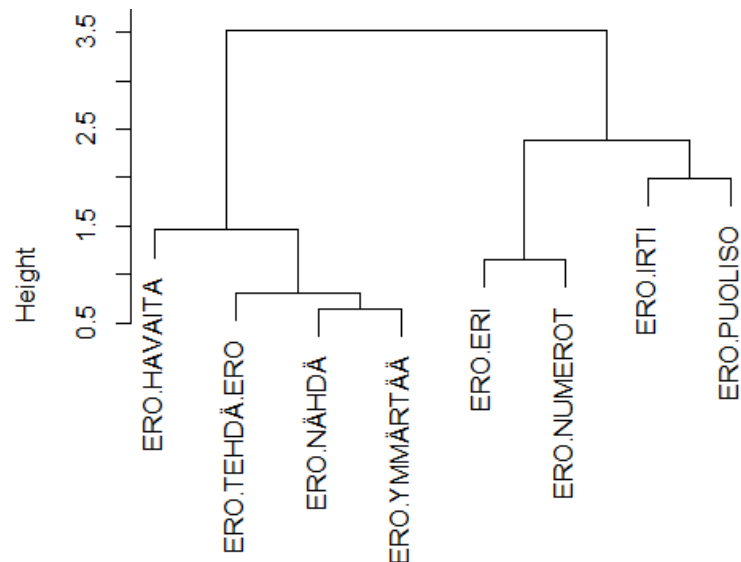
Transitiivilauseen objekteina esiintyviä skeemoja ovat ERO.HAVAITA, ERO.TEHDÄ.ERO (n=54) ja ERO.YMMÄRTÄÄ (n=61), joista viimeisin esiintyy lisäksi lauseobjektina (n=23).

Herkkyystasolla 1 skeemat ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ karsastavat luokkaa Y_MUOTO_NA (eli Y_MUOTO_LÄSNÄ, ERO.YMMÄRTÄÄ 66.7 %; ERO.TEHDÄ.ERO 72.1 %) ja suosivat luokkaa Z_MUOTO_NP (ERO.YMMÄRTÄÄ = 73.3 %; ERO.TEHDÄ.ERO 68.5 %). Ne siis suosivat Y- ja Z-argumentin ilmipanoa. Skeeman ERO.HAVAITA havainnot sen sijaan hajautuvat luokkiin Z_MUOTO_NA~NP (kummassakin 38.5 %), ja myös Y-argumentti on yleensä ilmipanematta (Y_MUOTO_NA = 65.4 %). Tässä se muistuttaa skeemoja ERO.NUMEROT, ERO.PUOLISO ja ERO.IRTI. Semanttisesti skeemojen Y-argumentit hakeutuvat luokkaan ELOTON (ERO.YMMÄRTÄÄ 60.4 %; ERO.TEHDÄ.ERO 72.3 %; ERO.HAVAITA 62.5 %). Myös Z-argumentit hakeutuvat luokkaan Z_SEM_ELOTON (ERO.YMMÄRTÄÄ 55.3 %; ERO.TEHDÄ.ERO 77.3 %) tai hajautuvat luokkien ELOTON ja KONKRETIA välille (ERO.HAVAITA, molemmissa 44.4 %). Esimerkissä *Vielä 40 vuotta sitten ymmärrettiin, mikä ero on kurinpidolla_{Y_SEM_ELOTON} ja pahoinpitelyllä_{Z_SEM_ELOTON}* (K1453) molemmat ovat elottomia. Muuttujan X_MUOTO erot menevät ”virhemarginaalin sisälle”: NP-agenttien rinnalla ERO.TEHDÄ.ERO suosii verbiin inkorporoituja ja persoonapronominiin spesifioituja agentteja (X_MUOTO_INKORP.+PERS = 33.4 %) ERO.HAVAITA nollapersonaalisia agentteja (X_MUOTO_NOLLA = 36 %). Skeemalla ERO.YMMÄRTÄÄ on suurin skeemansisäinen NP-osuus (X_MUOTO_NP = 41 %). Ilmipannut agentit ovat aina ihmistarkoituksia.

Tässä yhteydessä on todettava, että *ero*-lausekkeet skeemojen ERO.YMMÄRTÄÄ ja ERO.TEHDÄ.ERO sisällä käyttäytyvät usein samankaltaisesti kuin skeemassa ERO.ERI. Ero on siinä, että edelliset kuuluvat skeemoihin, jotka muodostuvat konstruktiosta ”transitiiviverbi + *ero*-nominin”, jälkimmäinen taas esiintyy ”paljaissa” kopula + *ero* -konstruktioissa. Siksi skeeman ERO.YMMÄRTÄÄ korpustoteumassa *Onko kirjoittaja ymmärtänyt, mikä ero on pedofiililla ja opettajalla, joka kiusaa ja lannistaa oppilastaan* (K1523) alleviivattu jakso on oikeastaan skeeman ERO.ERI toteuma. *ero*-nominin transitiiviset skeemat sisältävätkin osakomponenttina skeeman ERO.ERI, ’eroavaisuuden semeemin’, jota transitiiviverbi työstää. Toinen huomio on, että klusterianalyseissä ei ole mukana muuttujaa, joka erottaisi lauseobjektin paikalla olevat, epäsuoran kopulakysymyslauseen osana toimivat *ero*-lausekkeet (K1523) ”paljaana” objektitäydennyksenä toimivista *ero*-lausekkeista (ks. s. 51).

Lopuksi tarkastelen klusterianalyysin tuottamaa käsitystä *ero*-nominista klusterikuvion 7 avulla. Klusteroitavana ovat *ero*-nominin skeemat, joiden käyttäytymisprofileihin on syötetty muuttujien X_MUOTO, X_SEM, Y_MUOTO, Y_SEM, Z_MUOTO, Z_SEM, RAKENNE ja RESIPROOKKI alakategoriat samankaltaisine luokkayhdistelyine kuin khiin neliö -testeissä¹⁴⁹. Olen ottanut erilleen skeemaan ERO.HAVAITA pakatun skeeman ERO.NÄHDÄ.

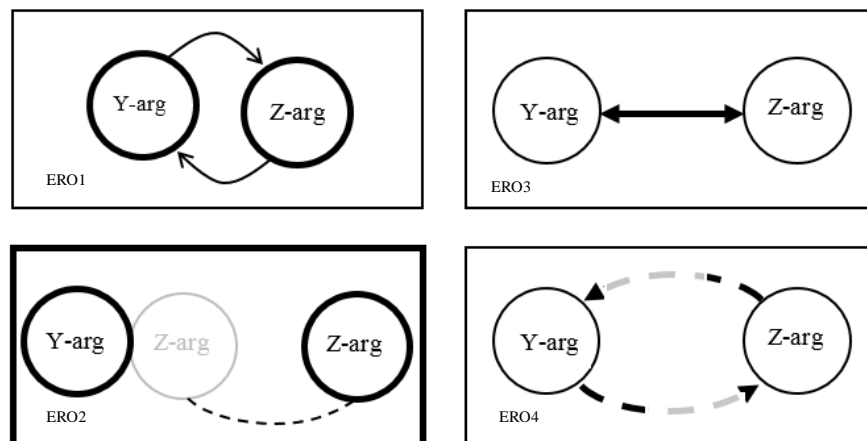
¹⁴⁹ Joitakin uudelleen järjestelyjä on tehty, kuten SEM-muuttujista on poistettu NA-luokat.



Kuvio 7. *ero*-nominin klusterianalyysi. 32 ID-merkkiä. Etäisyys: *euclidean*, Metodi: *Ward*.

Klusterointi tukee jo tehtyjä havaintoja. Ensimmäisenä yhdistyvät transitiiiviset skeemat ERO.NÄHDÄ ja ERO.YMMÄRTÄÄ, joihin yhdistyy skeema ERO.TEHDÄ.ERO. Nämä skeemat ovat kaikista samankaltaisimpia. Ylemmällä tasolla niihin yhdistyy skeema ERO.HAVAITA, mikä vihjaa, että se ja aikaisemmin yhdistyneet skeemat ovat suhteessa jonkin verran erilaisempia, joskin yhä aikailla samankaltaisia verrattuna vaikka skeemaan ERO.PUOLISO. Skeemat ERO.ERI ja ERO.NUMEROT muodostavat klusterin kolmantena. Skeema ERO.ERI on siis samankaltaisempi skeeman ERO.NUMEROT kanssa kuin esimerkiksi skeeman ERO.HAVAITA kanssa. Viimeiseksi yhdistyy klusteri, johon kuuluvat skeemat ERO.IRTI ja ERO.PUOLISO. Koska yhdistyminen tapahtuu suhteellisen myöhään, voidaan tulkita, että kaltaisuus on vähäisempää kuin esimerkiksi skeemojen ERO.ERI ja ERO.NUMEROT kaltaisuus.

Lopuksi esitän leinolais-langackerilaisen kuvion 7b (ks. Leino 1993: 75, 92; Langacker 1987: 68–69), joka hahmottaa *ero*-nominin ”konseptuaalisia jäsennystrategioita”.



Kuvio 7b. *ero*-nominin konseptuaalisia jäsennystapoja.

Kuvioissa 7b¹⁵⁰ laatikot viittaavat käsitteistykseen (ks. Leino mts. 72) Paksulla viivalla korostetut osiot viittaavat kognitiivisen kieliovin kuvio/tausta-jaon (ks. Leino mts. 46) nojalla predikaattien hahmoon tai kuvioon (siihen, mikä käsitteistyksessä on kohosteista) ja jäljelle jäävät elementit taustaan tai kehykseen (siihen, jota vasten kohosteinen osa hahmotetaan).

Kuviossa predikaatti ERO1 heijastelee skeemojen ERO.ERI, ERO.YMMÄRTÄÄ, ERO.TEHDÄ.ERO ja osin ERO.HAVAITA ilmoittamia käsitteistyskäsitteitä, joissa huomion keskipisteenä on kahden entiteetin – Y ja Z – argumenttien eroavaisuus (*Kissan*_{Y-arg.} ja *koiran*_{Z-arg.} *ero on se, että* [...]). Predikaatti ERO3 heijastaa skeeman ERO.NUMEROT ilmoittamaa käsitteistystä, jossa huomio kohdistuu kahden (usein ilmipanamattoman) entiteetin väliseen välimatkaan tai vastaavaan (*Lähijunalla Leppävaarasta Lempäälään on 28.73 € ja Pasilasta 20.77 €, eroa on 7.96 €, K1075*). Predikaatti ERO2 ilmoittaa skeeman ERO.IRTII hahmotettavan (suuntaisen) tilanteen (*Kunnat*_{Y-arg.} *suunnittelevat eroa Kuntaliitostaz-arg.* K1210). Se on analoginen *ero*-verbien konseptuaaliseen malliin, jossa X-argumentti (*erota*-verbissä Y-argumentti) aiheuttaa Y-argumentin SEPARAATIO:n Z-argumentista. Predikaatissa ERO2 korosteisena on itse käsitteistys (laatikko). Se vihjaa, ettei kyseessä *ero*-verbien tapaan ole prosessi vaan nonimaalistettu tilanne. Lisäksi korostettuna ovat Y- ja Z-argumentti(pallot), mikä vihjaa niiden, että ne on usein ilmipantuina lauseessa. Predikaatti ERO4 ilmoittaa, että kyse on entiteettien suhteen katkeamisesta (*Sitten kun ero tulee, vaimo mieheltä puolet omaisuudesta*, K1658). Kuviossa katkoviivanuolet viittaavat suhteeseen ja harmaat railot suhteen katkeamiseen. Korosteista on nimenmaan suhteen katkeaminen, ”ero”. Korostamattomat Y- ja Z-argumentti(pallot) vihjaavat täydennysten ilmipanamattomuutta.

Näiden konseptuaalisten jäsennystapojen voidaan ajatella periytyvän *erota*- ja *erottaa*-verbeille, joissa verbijohtimet lisäävät käsitteistykseen temporaalisen hahmon ja muuntelevat rooliargumenttien jäsentymistä (esimerkiksi lisäämällä kuvioon kolmannen pallon, jota voi luonnehtia AIHEUTTAJA:ksi tai KOKIJA:ksi, ks. s. 52). *erottaa*-verbissä Y- ja Z-argumenttien vaihtoehtoisten jäsennysstrategioiden RAKENNE_KONNEKTORI~REKTIO – eli opposition $ONP_{+ja+NP} (+RESIPROOKKI) : OA_{ero}$ (*erottaa kissat ja koirat [toisistaan] : erottaa kissat koirasta*) – taas voidaan nähdä heijastelevan predikaattien ERO1 ja ERO2 konseptuaalisia eroja. Nämä kysymykset ovat kuitenkin työn ulottumattomissa. Lopetan *ero*-nominin analyysin tähän. Uskon luvun 5 analyysien vastanneen tutkimuskysymyksiini skeemojen lekseemin-

¹⁵⁰ Diagrammikuvausta kritisoi Itkonen, jonka mukaan (1997: 71–72) ihmiset saattavat lukea yksinkertaisintakin kuviota täysin eri tavoin, sillä kuva ei aseta tulkinnalle rajoja eikä kerro, mikä on oikea, mikä väärä tulkinta.

sisäisistä suhteista. Yksi kysymys on silti vailla vahvistusta, nimittäin, millaisia lekseemirajat ylittäviä kytköksiä skeemoilla on. Tähän yritän vastata luvun 6 klusterianalyysissä. Ennen sitä kokoon ajatuksia yksimuuttujaisesta analyysivaiheesta.

5.5 Yksimuuttujaisen analyysin koonti

Luvun 5 yksimuuttujainen analyysi koostuu kolmesta osatutkimuksesta – *erota*-verbin (luku 5.2), *erottaa*-verbin (luku 5.3) ja *ero*-nominin (5.4) analyyseistä. Yksimuuttujainen analyysi osoittaa, että skeemojen välillä on semanttisia ja rakenteellisia eroja, jotka palautuvat X-, Y- ja Z-argumenteiksi nimeämieni rooliargumenttien ko(n)tekstuaaliseen käyttäytymiseen – tai *kollikatiivisiin profiileihin*. Kollikatiivisella profiililla (ks. luku 2.4) tarkoitan suunnilleen samaa kuin Kishner ja Gibbs (1997: 27–28), joiden hypoteesi on, että polyseemisen sana eri merkitykset korreloivat suhteessa rajattuun joukkoon ko(n)tekstuaalisia piirteitä (esimerkiksi verbin täydennyksiin), joita korpuslingvistiikassa kutsutaan kollikaatioiksi (Gries 2006: 63). Samankaltainen lähestymistapa on Jandalla ja Solovyellä (2009: 1–2), jotka käyttävät käsitettä *konstruktionaalinen profiili* (ks. s. 25 alav. 32) kuvaamaan (synonyymisten) nominien leksikaalisia kollikaatioita. Klusterianalyysin yhteydessä puhun lisäksi käyttäytymisprofiileista, jotka ovat alakategorioiden skeemakohtaiseen prosentijakaumaan pohjaavia vektoreita. Tulkintani on, että skeemoilla on paitsi yksilöllinen käyttäytymisprofiili myös yksilölliset kollikatiiviset profiilit, joita työssä kuvataan argumenttirooleihin pohjautuvilla rakennekuvauksilla. Profiilien todellinen luonne on kuitenkin rikkaampi kuin työni karkeistettu muuttujisto osoittaa. Ne voidaan nähdä tietorakenteiksi, konstruktioiksi, jotka sisältävät monenlaista semanttista, morfologista ja syntaktista tietoa (vrt. Goldberg 1995: 1, 6–7).

Oikeastaan voidaan teoretisoida kahdenlaisia profiileja: lekseemiskeemojen ja pesyeskeemojen profiileja, jotka ovat edellistä skemaattisempia. Esimerkki lekseemiskeemaan kytkeytyvästä kollikatiivisesta profiilista on Z-argumentin koodaaman lähteen eli rektioadverbiaalinen ilmipanemattomuus skeemoissa ERO.PUOLISO ja EROTA.PUOLISO, mikä erottaa niitä sellaisten skeemojen profiileista, joissa Z-argumentti on yleensä ilmipantu. Yhdessä skeemat ERO.PUOLISO ja EROTA.PUOLISO muodostavat hypoteettisen pesyeskeeman, 'parisuhteen päättymisen merkityksen', jonka profiiliin kuuluu lähteen ilmipanemattomuus sekä joukko muita ominaisuuksia, kuten, että Y-argumenttina on joukko- tai yksilötarkoitteinen olento.

Ennen profiilien esittämistä pysähdyn hetkeksi pohtimaan, miten muuttujat soveltuivat analyysin tarpeisiin – tai: mitkä muuttujista toimivat. Yhteenvedojen taulukoista 30, 40 ja 51

näky, että voimakkainta assosiaatio skeemojen ja yksittäisten luokkien välillä on argumenttien semanttista ontologiaa ilmoittavilla kategorioilla. *ero*-nomini tekee kuitenkin poikkeuksen. Verbikuvauksen pohjalta luomieni muuttujien soveltaminen *ero*-nominiin ei osoittautu täysin onnistuneeksi. Usein luokkajako täytyy supistaa dikotomiaan ELOTON : ELOLLINEN. Tosin usein jo läsnä : ei-läsnä (NP : NA) -jaottelu¹⁵¹ riittää tuomaan esiin skeemojen eroja. Muuttujan RAKENNE luokkajako soveltuu paremmin *ero*-verbien kuin *ero*-nominin tarkasteluun, joskin osoittautuu, että myös *ero*-nominilla (ERO.IRTI) on elatiivirektiota. *ero*-verbeillä jatkumo JOUKKO : YKSIÖ : INSTITUUTIO : ELOTON (ABSTRAKTI) : KONKRETIA tuottaa monia merkitseviä erotteluja. Myös argumenttien läsnä- tai poissaoloa mittaavat muuttujat toimivat suurimmalta osin hyvin. Harvinaisemmat luokat (kuten PRON, RAKENNE tai PERS) synnyttävät yksilöllisiä konfiguraatiota. Esimerkiksi skeeman EROTTAA.ERI profiiliin, joka ensinnäkin käsittää kaikkien argumenttien läsnäolon, kuuluu, että agentin spesifioi eloton tarkoite, joka usein on koodattu *se*-tukipronomini + lause -rakenteella (X_MUOTO_RAKENNE).

Pohdin muuttujiston onnistuneisuutta vielä luvussa 8 ja siirryn nyt profiilien tarkasteluun. Yksimuuttujaisen analyysin yksityiskohtaisuuden ja skeemojen lukumäärän vuoksi en selosta skeemojen luonnehdintoja uudestaan, sillä ne ovat esillä yhteenvedoissa, joissa on usein esitetty myös skeemakohtainen protoesimerkki. Taulukossa 52 formuloin kullekin skeemalle rakenneskeeman, joka heijastelee skeeman yksilöllistä kollikatiivisesta profiilia. Myös kollikatiivisen profiilin voi ymmärtää prototyyppiseksi kategoriaksi, jolla on eriasteisia sanktioitumia (vrt. luku 2.2.1). Rakenneskeemoja ei tulekaan ymmärtää ehdottomiksi ehtojen listauksiksi, sillä yleensä skeemat eivät esiinny pelkästään profiilin esittämissä luokissa. Profiilit ilmaisevat pikemminkin luokkien tendenssejä saman johdostason skeemojen (*erota*, *ero*, *erottaa*) välillä. Prototyyppinäkin ne ovat ehdotuksia.

Esittelen profiilien koodaustavan skeeman EROTA.IRTI rakenneprofiilin avulla. Se on ”Y.arg{NP~PERS_**yksilö~instituutio**} +REKTIO~VAIN_Y+ Z.arg{NP~NA_**instituutio**}”. Lausekkeet Y.arg{} ja Z.arg{} kertovat mistä argumentista on kyse. Kaarisulkujen sisällä tulee ensiksi MUOTO-kategoria, jota _-merkin jälkeen seuraa SEM-kategoria. +-merkkien väliin jäävä jakso viittaa muuttujaan RAKENNE. Jos lopussa on [+R], se tarkoittaa, että skeema sallii resiprookin. Tummennetut luokat tarkoittavat herkkyystasolla 3 positiivista merkitsevyyttä kulloisessakin jakaumataulussa, tummentamattomat herkkyystasoa 1. Edelliset ilmaisevat erityisen poikkeavaa käytöstä. Mikäli herkkyystason 3 tai 1 tuloksia ei ole, on esillä skeemakohtaisesti suurin prosentuaalinen luokka, joka on lausekkeessa myös silloin, kun skeemalla on herkkyystason 3 tai 1 tulos, joka ei kuitenkaan ole prosentuaalisesti suurin luokka.

¹⁵¹ Työssänimitatuista X-, Y- ja Z-argumenteista ilmipantuja on 2692 ja ilmipanemattomia 2249 (45.5 %).

Herkkyystasot eivät ole vertailukelpoisia edes saman skeeman eri argumenttien kesken, sillä ne ovat taulukko-kohtaisia. Koska herkkyystason 1 havainnot ovat tummentamatta, luokitus ei tee eroa isoimman skeemakohtaisen prosentiosuuden ja herkkyystason 1 havainnon välillä. Aaltoviiva eli ”~”-merkki tarkoittaa luokkien välistä vaihtelua, ja milloin luokkia on monta, ensimmäisenä on se, jolla on isompi jakauma. Lauseke ”Y.arg{NP~PERS_yksilö~instituutio}” tarkoittaa siten, että skeeman EROTA.IRTI esiintymissä Y-argumentit hakeutuvat luokkiin Y_MUOTO_NP ja Y_MUOTO_PERS. Lisäksi skeema hakeutuu herkkyystasolla 3 luokkiin Y_SEM_YKSIÖ ja Y_SEM_INSTITUUTIO. Lauseke ”+REKTIO~VAIN_Y +” tarkoittaa, että skeema hakeutuu herkkyystasolla 3 luokkaan RAKENNE_VAIN_Y. Prosenttijakauman nojalla sen yleisin luokka on kuitenkin REKTIO.

Skeemojen ryhmittelyyn vaikuttavat yksimuuttujaisesta analyysistä nousseiden havaintojen lisäksi hypoteesini siitä, että pesyeestä on abstrahoitavissa pesyeskeemoja, esimerkiksi ’laadun ja ominaisuuden eroavaisuuden skeema’. Taulukko 52 toimii samalla koontina yksimuuttujaisen vaiheen tuloksista.

Taulukko 52. *erota-* ja *erottaa-*verbin kollikatiivisten profiilien rakenneskeemat.

Skeema	Profililauseke: X.arg{ } Y.arg { } +RAKENNE+ Z.arg { } [+R]
HAVAITA1	X.arg {NP~NOLLA~PERS_eloll.} Y.arg {NP~PRON.RAK_eloton~konkr.} +VAIN_Y+ Z.arg {NP_eloton}
HAVAITA2	X.arg {NP~NOLLA~PERS_eloll.} Y.arg {NP~PRON.RAK_eloton~konkr.} +REKT.~KONN.+ Z.arg {NP~PRON.RAK_eloton} [+R]
LUOKITTELU	X.arg {NOLLA_eloll.} Y.arg {NP_eloton} +KONNEK.+ Z.arg {NP_eloton} [+R]
ERO.YMMÄRTÄÄ	X.arg {NP_eloll.} Y.arg {NP_eloton} +KONNEK.+ Z.arg {NP_eloton} [+R]
ERO.TEHDÄ.ERO	X.arg {PERS_eloll.} Y.arg {NP_eloton} +KONNEK.+ Z.arg {NP_eloton} [+R]
ERO.HAVAITA	X.arg {NOLLA_eloll.} Y.arg {NP_eloton} +Y_JAOLLINEN+ Z.arg {NA_eloton} [+R]
ERO.NUMEROT	X.arg {NA} Y.arg {NA_eloton} +NA+ Z.arg {NA} [+R]
EROTA.ERI	Y.arg {NP~PRON_eloton~konkr.} +REKT.~KONN.+ Z.arg {NP~PRON_eloton~eloll.} [+R]
ERO.ERI	X.arg {NA_eloton} Y.arg {NP~PRON.RAK_eloton} +KONNEK.+ Z.arg {NP_eloton} [+R]
EROTTAA.ERI	X.arg {NP~PRON.RAK_eloton} Y.arg {NP_joukko} +REKT.+ Z.arg {NP~PRON.RAK_eloll.}
EROTTAA.ABSTR.	X.arg {NA_eloton} Y.arg {NP_eloton} +REKTIO+ Z.arg {NP_eloton} [+R]
EROTTAA.KONKR.	X.arg {NA_elollinen} Y.arg {NP_konkretia} +REKTIO+ Z.arg {NP_konkretia} [+R]
EROTTAA.IRTI2	X.arg {NA} Y.arg {NP_instit.} +REKTIO~KONNEKTORI+ Z.arg {NP_instit.} [+R]
EROTTAA.IRTI	X.arg {NA_eloll.} Y.arg {NP_yksilö~joukko} +VAIN_Y+ Z.arg {NA_instituutio}
ERO.IRTI	X.arg {NA} Y.arg {NP_elollinen} +REKTIO~VAIN_Y+ Z.arg {NA_elollinen}
EROTA.IRTI	Y.arg {NP~PERS_yksilö~instituutio} +REKTIO~VAIN_Y+ Z.arg {NP~NA_instituutio}
EROTA.KIRKKO	Y.arg {NP~NOLLA~INKORP._yksilö~joukko} +REKT.~VAIN_Z+ Z.arg {NP_instit.}
EROTA.PUOLISO	Y.arg {NP~NOLLA_joukko} +Y_JAOLLINEN~NA+ Z.arg {NA_elollinen}
ERO.PUOLISO	X.arg {NA} Y.arg {NA_elollinen} +NA+ Z.arg {NA}

Jos monimutkaiselta vaikuttavaan taulukkoon syventyy, sieltä paljastuu kiinnostavia yhteyksiä. Esimerkiksi skeemojen EROTTAA.HAVAITA1, EROTTAA.HAVAITA2 ja EROTTAA.LUOKITTELU agentteina esiintyy elollistarkoitteisia NP-agentteja, persoonapronomineja sekä nollapersoonassa olevaa agenttia. Y- ja Z-argumentteina esiintyy elottomia tarkoitteita sekä lausetäydennyksiä (PRON.RAK, PRON + RAKENNE). Elollisia NP-agentteja sekä persoonapronominissa tai nollapersoonassa esiintyviä agentteja esiintyy myös skeemoilla ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ ja ERO.HAVAITA, joiden Y- ja Z-argumentteina niin ikään esiintyy elottomia tarkoitteita. Skeemojen EROTA.PUOLISO ja ERO.PUOLISO profiiliin kuuluu, että Y-argumentti ei välttämättä ole ilmipantu, ja milloin on ilmipantu, se on joukko- tai yksilötarkoitteinen. Z-argumentti jää ilmipane-matta (Z_MUOTO_NA). Skeemojen ERO.ERI, EROTA.ERI ja EROTTAA.ERI profiiliin kuuluu, että ilmipantu Y-argumentti on elollinen, ja usein nimenomaan yksilötarkoitteinen (Y_SEM_YKSIÖ). Z-argumentti jää usein ilmipane-matta, mutta milloin se on ilmipantu, se on monesti luokan INSTITUUTIO-tarkoite. En jää tutkimaan taulukkoa tämän enempää, vaan siirryn kohti monimuuttujaista analyysiä.

Yksi yksimuuttujaisesta analyysistä kumpuava menetelmällinen kysymys vaatii kuitenkin vielä pohdintaa. Kysymys kuuluu, ovatko luvun 5 yksityiskohtaiset *post hoc* -testaukset tarpeellisia. Useinhan skeemojen ja luokkien väliset painotussuhteet paljastuvat jakaumaa tarkkaamalla tai viimeistään skeemakohtaisia ja skeemojen välisiä prosenttijakaumia tarkastelemalla. Taulukko 52 antaa vihjeen siitä, mitä järkeä *post hoc* -testauksilla on. Hyöty on siinä, että ihmisen on sittenkin hankala havainnoida skeemojen välisiä painotussuhteita, koska yhtä aikaa pitäisi tarkata skeemojen välistä jakaumaa ja skeemojen sisäistä jakaumaa sekä lisäksi sitä, onko luokissa riittävästi havaintoja, jotta voidaan tehdä relevantteja päätelmiä. Usein *post hoc*-testaukset paljastavat kytköksiä, jotka jäisivät ihmissilmältä havaitsematta. Toisaalta ne voivat myös kyseenalaistaa ihmissilmän tekemiä päätelmiä. On muistettava, että *post hoc* -mittaustulokset ovat taulukkokohtaisia (”muuttujasidonnaisia”), ja mikäli halutaan tarkata muuttujien välisiä assosiaatiosuhteita, tarvitaan mittareita kuten Cramérin *V* (symmetrinen), λ :ta ja *U*:ta (epäsymmetriset). Niillä ei toisaalta voida tarkastella skeemojen luokkien yksittäisiä kytköksiä, minkä takia olen sivuuttanut ne koonnista. Kolliktiivisten profiilien yksi uskottavuusongelma on se, että ne on koostettu toisistaan eristyksissä tarkasteltujen yksimuuttujaisten testausten pohjalta. Niiden rakenne voi muodostua erilaiseksi, mikäli luokkia tutkittaisiin monimuuttujaisesti vuorovaikutuksessa toisiinsa.

Lopuksi esitän taulun 53, joka on samalla yhteenvedo semanttisten rooliargumenttinen

vaihtelusta ja yksinkertaistettu havainnointi klusterianalyysin mekaniikasta¹⁵². Tauluun on koottu X-, Y- ja Z-argumentteja koskevista SEM-muuttujista skeemakohtaisista suurimmat ja toiseksi suurimmat prosenttiarvot. Taulu koostuu neljästä sarakkeesta, josta ensimmäinen esittää muuttujaluokan, jota vasten *ero*-skeemat on ristiintaulukoitu ilman etutunnusta (EROTA, ERO ja EROTTAA). Musta fontti signaloi, että luokalla on skeeman sisäisessä prosenttijakaumassa valta-asema, harmaa fontti signaloi toiseksi suurinta osuutta. Jotta analogiat erottuisivat, lihavoituna ovat skeemat, jotka ilmoittavat kognitiivista toimintaa ja aistiha-
vaintoa eli skeemat EROTTAA.LUOKITTELU, EROTTAA.HAVAITA1–2, ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ ja ERO.HAVAITA. Alleviivattuna ovat skeemat, joiden lopputunniste on ERI, ja kursivoituina skeemat, joiden lopputunniste on PUOLISO. Taulukko toimii johdantona klusterianalyysin siksi, että muuttujakohtaiset prosenttijakaumat ovat täsmälleen samoja arvoja, joista käyttäytymisprofiilit muodostuvat (ks. luku 2.4) ja joiden pohjalta klusterointialgoritmi klusteroi (muunnettuaan käyttäytymisprofiilit ensin samankaltaisuustaulukoksi).

Taulukko 53. Muuttujien X_SEM, Y_SEM ja Z_SEM jakauma¹⁵³.

ID-merkki	erota-verbi	ero-nomini	erottaa-verbi
Y_SEM_JOUKKO	PUOLISO, KIRKKO	PUOLISO	ERI, ABSTRAKTI, LUOKITTELU, IRTI, KONKRETIA
Y_SEM_YKSIÖ	IRTI, KIRKKO, PUOLISO	IRTI	IRTI, ERI
Y_SEM_INSTITUUTIO	IRTI	IRTI	IRTI2,
Y_SEM_ELOTON	ERI	YMMÄRTÄÄ, TEHDÄ.ERO, HAVAITA, ERI, NUMEROT	LUOKITTELU, HAVAITA2, HAVAITA1, ABSTRAKTI, IRTI2
Y_SEM_KONKRETIA	ERI	YMMÄRTÄÄ	KONKRETIA, HAVAITA1, HAVAITA2
Y_SEM_NA		NUMEROT, PUOLISO, TEHDÄ.ERO, HAVAITA, ERI	
Z_SEM_JOUKKO			ERI,
Z_SEM_YKSIÖ	PUOLISO		ERI, KONKRETIA, HAVAITA1
Z_SEM_INSTITUUTIO	IRTI, KIRKKO	IRTI	IRTI2, IRTI
Z_SEM_ELOTON	ERI	TEHDÄ.ERO, ERO.YMMÄRTÄÄ, NUMEROT, HAVAITA, ERI	LUOKITTELU, HAVAITA2, ABSTRAKTI
Z_SEM_KONKRETIA			KONKRETIA
Z_SEM_NA	PUOLISO, KIRKKO, IRTI, ERI	ERI, NUMEROT, IRTI, PUOLISO, HAVAITA, YMMÄRTÄÄ, TEHDÄ.ERO	HAVAITA1, IRTI, ABSTRAKTI, IRTI2, LUOKITTELU, HAVAITA2
X_SEM_JOUKKO			LUOKITTELU
X_SEM_YKSIÖ		YMMÄRTÄÄ, TEHDÄ.ERO, HAVAITA	KONKRETIA, HAVAITA1, HAVAITA2
X_SEM_INSTITUUTIO			IRTI
X_SEM_ELOTON		ERI	ERI, ABSTRAKTI
X_SEM_KONKRETIA			
X_SEM_NA	PUOLISO, KIRKKO, IRTI, ERI	ERI, IRTI, PUOLISO, NUMEROT, TEHDÄ.ERO, YMMÄRTÄÄ, HAVAITA	ABSTRAKTI, IRTI, IRTI2, KONKRETIA, LUOKITTELU, HAVAITA1, HAVAITA2, ERI

¹⁵² Esitys on karkeistus, sillä käyttäytymisprofiileissa olevien ID-merkkien arvot lasketaan niin, että kaikki alakategoriat ovat edustettuina ja siten, että kukin alakategoria saa täsmällisen prosentuaalisen tunnusluvun.

¹⁵³ Joillakin luokilla on esiintymä vain yhdessä luokassa (jossa on 100 % osuus), kuten skeemalla ERO.PUOLISO suhteessa muuttuun Z_SEM. Siksi sille ei ole merkitty toiseksi sijoittuvaa harmaata arvoa.

Silmäilemällä taulukkoa havaitaan, että kognitiivista toimintaa ja aistihavaintoa koodaavat skeemat EROTTAA.HAVAITA2 ja ERO.YMMÄRTÄÄ korreloivat kuusi kertaa, kun sivuutetaan se, onko kyseessä suurin vai toiseksi suurin skeemakohtainen jakauma. Jos käyttäytymisprofiileihin pohjautuvaa lähestymistapaa mukaillen assosiatiivisuuden laukaisee suunnilleen yhtä iso prosenttiosuus, skeemoilla on neljä yhteistä korrelaatiota (kaksi harmaata ja kaksi mustaa). Skeemat ERO.ERI ja EROTA.ERI korreloivat neljästi, ja skeemat ERO.PUOLISO ja EROTA.PUOLISO kolmesti. Skeemat ERO.IRTI, EROTA.IRTI ja EROTTAA.IRTI korreloivat täydellisesti neljästi ja skeemat ERO.IRTI ja EROTA.IRTI viidesti.

Lopetan yksimuuttujaisen analyysin yhteenvedon tähän ja siirryn monimuuttujaiseen analyysivaiheeseen, jonka tarkoitus on antaa lisätietoa skeemojen välisistä suhteista.

6 Monimuuttujainen analyysi ja sen tulokset

Jatkan *ero*-skeemojen analyysiä monimuuttujaisin menetelmin. Luvussa 6.1 tutustun käyttämäni analyysityökalujen teoriaan. Luvussa 6.2 tutkin *ero*-pesyettä kahden eri samankaltaisuusmittarin, *euklidisella* etäisyyden (A-klusterointi, luku 6.2.1) ja *canberra*n (B-klusterointi, luku 6.2.2), tuottamien klusteristruktuurien avulla. Klusterien välisiä suhteita mitaan *t*-arvoilla (ks. Divjak & Gries 2008). Luvussa 6.2.3 teen yhteenvedon monimuuttujaisen analyysin tuloksista. Luvun monimuuttujaisen analyysin tarkoitus on etenkin tarjota näkökulmia *ero*-pesyeen jatkotutkimusmahdollisuuksiin. Samalla se antaa alustavia havaintoja siitä, millaisia kytköksiä *ero*-skeemojen välillä muuttujasysteemiini valossa on. Analyysi on yksimuuttujaisen analyysin yksityiskohtaisuuteen verrattuna suhteellisen pinnallista.

6.1 Agglomeratiivinen hierakkinen klusterianalyysi

Yksimuuttujaisen analyysin ongelma on, että se keskittyy yksinomaan yksittäisten piirteiden tarkasteluun, kun taas monimuuttujaisessa analyysissä voidaan tarkastella lukuisten muuttujien yhteisvaikutusta tutkitussa ilmiössä (Arppe 2008: 113). Käyttämäni monimuuttujainen analyysimenetelmä on hierarkkinen klusterianalyysi. Se on kattokäsite tekniikoille, jotka identifioivat datasta muuttujajoukon suhteen samankaltaisia ryhmiä ja visualisoivat tulokset dendrogrammikuviiksi (ks. Baayen 2008: 138). Työssäni käytetty menetelmä on nimeltään agglomeratiivinen klusterianalyysi (jatkossa HAC), joka tarkoittaa alhaalta ylöspäin etenevää kokoavaa yhdistämistä. Semantiikan tutkimuksissa HAC-analyysiä käyttävät muun

muassa Divjak ja Gries (2006, 2009), Arppe (2008), Deshors (2010) ja Gries ja Otani (2010). Tiedossani ei ole, että klusterianalyysiä olisi sovellettu kognitiivisen kielentutkimuksen piirissä suomen kielen johto-oppiin.

On syytä korostaa, että datan ryhmittämis- ja visualisointimenetelmänä klusterianalyysi on kokeellinen (*an exploratory data analysis; a hypothesis-generating method*) (Gries 2012: 64; Manning & Schütze 1999: 497). Tämä näkyy esimerkiksi siinä, ettei klusteroitavalle aineistolle tai muuttujille esitetä kirjallisuudessa kummoisia tilastotieteellisiä vaatimuksia (toisin kuin esimerkiksi regressioanalyysissä, ks. Gries 2012: 75–76). Klusterointi ei toisaalta anna absoluuttisia tuloksia, vaan pikemminkin vihjeen siitä, millainen tutkimuskohteen struktuuri ehkä on. On tavallista, että klusterointia käytetään esitutkimuksessa (esimerkiksi luokkien lukumäärän sieventämisessä tai aineiston visuaalisen kokonaiskäsityksen haalimisessa) tai apuvälineenä muiden monimuuttujaisten keinojen, kuten regressioanalyysin rinnalla (näin esimerkiksi Deshorsilla 2010). (Manning & Schütze 1999: 500.)

HAC-analyysissä lähtökohtana on, että jokainen skeema edustaa lähtötilanteessa omaa klusteriaan. Analyysi etenee visuaalisesti katsottuna alhaalta ylöspäin kimputtamalla yhteen ne klusterit (skeemat), jotka osoittavat suurinta sisäistä samankaltaisuutta (*intra-cluster similarity*) ja erilleen ne klusterit, jotka osoittavat pientä keskinäistä samankaltaisuutta (*inter-cluster similarity*), kunnes kaikki klusterit on kimputettu yhdeksi klusteriksi (ks. Arppe 2008: 23). Mitä ylemmälle hiearkiasolalle klusterointi etenee, sitä suuremmiksi klusterit muodostuvat ja sitä ”skemaattisempaa” samankaltaisuus on. Tämä tarkoittaa sitä, että elementit, jotka yhdistyvät aikaisemmin, ovat samankaltaisia ja elementit, jotka yhdistyvät myöhemmin, ovat pikemminkin erilaisia (joskin yhä suhteellisen samankaltaisia, koska kyseessä on joukko samakantaisten sanojen ”eri merkityksiä”). Lisäksi kaavakuva antaa vihjeen siitä, kuinka itsenäisiä klusterit ovat: mitä pitempi välimatka klustereilla on, sitä itsenäisempi aiemmin klusteroitu skeema on suhteessa siihen myöhemmin yhdistyvään skeemaan. (Manning & Schütze 1999: 498–450; Divjak & Gries 2006: 38.)¹⁵⁴

HAC-analyysi turvautuu dendrogrammipuun luonnissa kahteen prosessiin, jotka ovat samankaltaisuuden mittaaminen ja elementtien yhteen niputtaminen. Samankaltaisuutta mitataan samankaltaisuustaulukolla (*similarity matrix*). Se lasketaan tutkijan valitsemalla samankaltaisuusmittarilla (*similarity metric*) vektoreista, joita lähestymistavassani kutsutaan käyttäytymisprofileiksi (ks. luku 2.4). On tutkijan valittavissa, mitä ID-merkkejä käyttäyty-

¹⁵⁴ Kognitiivisen kielentutkimuksen näkökulmasta voidaan tulkita, että klusterianalyysi heijastelee skeemojen välistä elaboraatioetäisyyttä (ks. Langacker 1987: 379; Divjak 2010: 107) kimputtamalla yhteen (ensiksi) ne skeemat, joiden keskinäinen elaboraatioetäisyys on pienintä.

misprofileihin kulloinkin sisällytetään. Elementtien yhteen niputtaminen eli dendrogrammipuun tuottaminen syntyy samankaltaisuustaulukon pohjalta yhdistämissäännön (*amalgamation rule*) avulla. Yhdistämissääntö määrittää algoritmit sille, milloin kaksi klusteria ovat tarpeeksi samanlaisia yhteen niputettavaksi ja milloin eivät. (Divjak & Gries 2006: 38–39.)

Sekä samankaltaisuusmittariksi että yhdistämissäännöksi on olemassa lukuisia vaihtoehtoja, eikä ole selvää, mikä niistä on paras (Arppe 2008: 23, alv. 20; Divjak & Gries 2006: 37). Sillä, mitä asetuksia käytetään, on kuitenkin huomattava vaikutus klusteroinnin tuloksiin (Baayen 2008: 138). Tyypillistä on, että aineistoon kokeillaan useita asetuksia ja valitaan ne, jotka tuottavat hyödyllisimmät tulokset (näin toimivat Arppe 2008 sekä Divjak & Gries 2006). Gries (2012; ks. myös Manning ja Schütze 1999: 500) rohkaisee tähän toimintamalliin, sillä eri tutkimusasetelmat edellyttävät erilaisia asetuksia. Tässä ongelmana on, että valitaan asetukset, jotka tuottavat haluttuja tuloksia, eikä raportoida lukuisia epäonnistuneita kokeiluja. Lukemissani kielitieteellisissä töissä suosituin yhdistämissääntö on *Ward*, jossa klusterien etäisyyttä mitataan sisäisen hajonnan avulla (*Ward* on käytössä Arppella 2008 ja kaikissa Griessin [ja Divjakin] työssä). Yhdistettävät klusterit valitaan siten, että yhdistäminen johtaa mahdollisimman pieneen lisäykseen klusterien sisäisessä varianssissa (Divjak 2010: 133). Sillä on taipumus luoda pienikokoisia klustereita (Divjak & Gries 2006: 37), mikä sopii polysemian tutkimiseen. Käytän HAC-analyyseissä yhdistämissääntönä *Wardia*.

Kognitiivisesti orientoituneissa kielitieteellisissä tutkimuksissa (esim. Divjak & Gries 2006, 2009; Deshors 2010) yleisin samankaltaisuusmittari lienee *canberra*. Se on erityisen toimiva, kun data sisältää paljon tyhjiä soluja (Divjak 2010: 132). Suoritan tässä tutkimuksessa kaksi johdospesyeen klusterianalyysii, joista ensimmäisessä käytän samankaltaisuusmittarina euklidista etäisyyttä (*euclidean distance*). Se on Divjakin ja Fiellerin (2014: 417) mukaan samankaltaisuusmittareista kaikista yksinkertaisin. Sitä käyttävät esimerkiksi Glynn (2014) ja Arppe (2008: 315). Toisen testin suoritan *canberralla*. Muita samankaltaisuusmittareita en käytä. Esitutkimuksesta ilmenneiden havaintojeni nojalla mittarien erot korostuvat, mitä vähemmän ID-merkkejä klusterointiin syötetään.

Samankaltaisuus- ja yhdistämissäännön lisäksi tutkijan tehtäväksi jää valita, kuinka monta klusteria dendrogrammin kulloinkin oletetaan sisältävän, mikä onkin tulosten tulkinnan keskeisin kysymys. Divjak ja Gries (2006: 39; ks. myös 2008) – joiden työt ovat ”klusterianalyysivetoisia” vastakohtana työlleni, jossa käytän HAC-analyysii lähinnä avauksena jatkotutkimusmahdollisuuksiin – pohtivat, ettei klusterien lukumäärän arvioinnin pitäisi

jääda introspektion ja ailahteleavan intuition varaan. Baayenin (2008: 142) suhde on leppä: “You first have to decide how many clusters **you think you need**¹⁵⁵ [...] Here, we opt for five clusters”. Myös Arppe (2008: 248) esittää HAC:n etuihin, ettei klusterilukua tarvitse päättää etukäteen, vaan sen voi sovittaa käytännön tarpeisiin, jotka Arppella (mts. 246. alav. 99) ovat kokeelliset: “I have decided to opt for preliminary, tentative treatment of the question with exploratory character, which may be refined in later research.”¹⁵⁶

Klusteriluvun kvantitatiiviseen määrittämiseen on joukko keinoja, muttei vakiintunutta strategiaa. Tutkija voi muun muassa tarkastella i) siluettikuviota tai selvittää luokituksen tarkkuutta joko ii) PAM-testillä (*a partitioning-around-medoids approach*) iii) tai *kmeans*-testillä, joka on konseptuaalisesti yksinkertaisin keino määrittää klusteriluku (Manning & Schütze 1999: 500). Omaksun työssäni hypoteettisen, arppemais-baayenilaisen lähestymistavan ja suosin klusterijakoja, jotka ovat semanttisesti mielekkäitä. Tarkoitukseni ei ole esittää tilastollisesti ankaran falsifioituja tuloksia, vaan pikemminkin avata näköaloja ero-pesyeeseen sen monimuuttujaista jatkotutkimusta silmällä pitäen.

Ennen klusterianalyysiä on vielä tarpeen selvittää yksimuuttujaisen ja monimuuttujaisen analyysin välistä liittoa. Gries (2003) jakaa empiirisen kielentutkimuksen testausmenetelmät kolmeen tyyppiin, joita Arppe (2008: 17) nimittää yksimuuttujaiseksi (*a univariate stage*), kaksimuuttujaiseksi (*a bivariate stage*) ja monimuuttujaiseksi (*a multivariate stage*) vaiheeksi. Nämä vaiheet eivät ole toisistaan erillisiä, vaan sekä yksimuuttujainen että kaksimuuttujainen vaihe alustavat monimuuttujaista vaihetta (Arppella logistista regressioanalyysiä). Yksimuuttujaisen vaiheen tarkoitus on etsiä ilmiön kannalta relevantit muuttujat ja selvittää näiden muuttujien kytköksen luonne (ks. Arppe 2008: 73), kun taas kaksimuuttujaisen vaiheen tarkoitus on selvittää, onko muuttujissa voimakkaasti korreloivia pareja eli muuttujia, joiden läsnäolo (*occurrence*) tai poissaolo (*nonoccurrence*) kytkeytyy jonkin toisen muuttujan läsnäoloon tai poissaoloon (Arppe mts. 105–106). Monimuuttujaisen vaiheen valmisteluissa on Arppen (mts. 116–118) mukaan kiinnitettävä erityistä huomiota muuttujapareihin, joissa kaksimuuttujaisessa analyysissä on havaittu huomattavaa positiivista tai negatiivista kytköstä¹⁵⁷. Tämä indikoi kollineaarisuutta (*collinearity*) tai multikollineaarisuutta (*multicollinearity*). Täydellinen multikollineaarisuus on merkki siitä, että jokin muuttuja on edustettuna kaksi kertaa regressioyhtälössä. Baayenin (2008: 181) mukaan vahva korrelaatio voi synnyttää monia ongelmia. Yksi Arppen (mts. 117) nostama syy korreloivien muuttujien

¹⁵⁵ Lihavointia ei ole alkuperäisessä lähteessä.

¹⁵⁶ Lähes tyhjentävän listauksen erilaisista klusteriluvun määrittämisen keinoista antaa Ybungalobill (2012).

¹⁵⁷ Kaksimuuttujaisessa vaiheessa saadaan esille esimerkiksi kielisysteemin päällekkäisyyksia, kuten että fi-niittiverbin ilmirakenteessa esiintyvä agentti pakottaa verbin päälukaksi aina aktiivin (Arppe 2008: 105–106).

poistamiseen on, etteivät ne lisää logistisen regressionmallin tieteellistä selitysvoimaa, vaan pikemminkin varastavat tilaa muuttujilta, jotka voisivat olla selittävämpiä.

Kirjallisuudessa klusterianalyysille ei yleensä esitetä rajoituksia muuttujien maksimilukumäärässä (vrt. *overfitting* eli *ylisovittaminen* regressioanalyysissä, Arppe mts. 116), eikä vaateita multikollinearisuuden välttämiseksi (ks. esim. Glynn 2012)¹⁵⁸. Klusterianalyysi ei osaa tehdä eroa relevantin ja irrelevantin muuttujan välillä, minkä takia siihen syötettyjen muuttujien tarpeellisuus täytyy olla teoreettisesti punnittua. Toisaalta ilmiön kannalta relevanttien muuttujien puuttuminen vääristää tuloksia niin kuin regressioanalyysissäkin (vrt. Arppe 2008: 116–119). Siksi sekä yksimuuttujainen että kaksimuuttujainen vaihe ovat paitsi tarpeellisia myös välttämättömiä klusteroinnin esivalmisteluissa (vrt. Arppe 2008: 104, 119). Työni laajuuden vuoksi en kuitenkaan suorita erillistä kaksimuuttujaisista vaihetta.

6.2 ero-pesyeen klusterianalyysi

Tässä luvussa suoritettujen klusterianalyysien pohjana oleviin käyttäytymisprofileihin (ks. luku 2.4) on sisällytetty kaikki luvussa 2.4 listatut ja luvussa 4 määritellyt muuttujat muuttujaa RAKENNE lukuun ottamatta, sillä havaintojeni perusteella sen ja muuttujien Y_SEM~MUOTO ja Z_SEM~MUOTO¹⁵⁹ välillä vallitsee erityisen voimakas assosiaatio. Luokan poistamisen vaikutukset klusteroinnin tuloksiin ovat kuitenkin vähäiset sekä euklidisella että canberralla suoritetuissa analyyseissä¹⁶⁰. Lopullisia muuttujia on 19, jotka sulussa olevine selityksineen ovat¹⁶¹: X_MUOTO, X_SEM (transitiivilauseen agentti), Y_MUOTO, Y_SEM (patientti tai intransitiivilauseen subjekti), Z_MUOTO, Z_SEM (lähde), RESIPROOKKI (*toisistaan, välillä* yms.), PÄÄLUOKKA (aktiivi, passiivi), LUOKKA (intransitiivinen, transitiivinen, kopula), KIELTO (kielteinen, neutraali), MODUS (indikatiivi, muu), TEMPUS (preesens, perfekti, muu), VERBI_RAKENNE (ero.verbi, verbiketju, muu, na), MODUS1 (kyky, na), MODUS2 (välttämättömyys, mahdollisuus, na), SYY (syy, na), AIKA (aika, na), VASTAUS (vastaus, na) ja

¹⁵⁸ Ketchen ja Shook (1996: 444) kirjoittavat multikollinearisuuden mahdollisista ongelmista klusterianalyysissä seuraavasti: ”High correlation among clustering variables can be problematic because it may **overweight one or more underlying constructs**. Thus, researchers may want to correct multicollinearity, especially if it is desirable that constructs be equally weighted.” Lihavointia ei ole alkuperäisessä lähteessä.

¹⁵⁹ Myös muuttujien Y_SEM ~ Y_MUOTO, X_SEM ~ X_MUOTO ja Z_SEM ~ Z_MUOTO välillä vallitsee vahva assosiaatio, minkä vuoksi poistin SEM-muuttujista alakategoriat NA, jotka korreloivat täydellisesti MUOTO-muuttujien NA-alakategorioiden kanssa. Tällaisia yhteyksiä kaksimuuttujainen analyysivaihe paljastaa. Muita kytköksiä ovat luokkien X_MUOTO_PERS kytkös (100 %) luokkiin X_SEM_JOUKKO ja X_SEM_YKSILÖ tai luokan X_MUOTO_RAKENNE~PRON kytkös (100 %) luokkaan X_MUOTO_NP. En kuitenkaan poistanut näitä luokkia.

¹⁶⁰ Muuttujan RAKENNE luokista ainoastaan luokka VAIN_Z sai esitutkimuksessa korkean *t*-arvon.

¹⁶¹ ID-merkkien lukumäärä vaihtelee tutkimuksesta toiseen (ks. tarkemmin Gries 2006, 2012).

KVANTTORI (kvanttori). Koska klusterianalyysi on herkkä poikkeaville havainnoille, yhdistin pienifrekvenssisii luokkia isommiksi ryhmiksi yksimuuttujaisen analyysin tulosten nojalla.

Yhteensä ID-merkkejä eli alakategorioita on 53¹⁶². Ennustettavia muuttujia eli skeemoja on yhteensä 20 ja korpusesimerkkejä (skeemojen EROTA.ABSTRAKTI ja EROTA.KONKRETIA poistamisen jälkeen) 1661. Niistä *ero*-skeemoja on 8 (ERO.IRTI, ERO.PUOLISO, ERO.TEHDÄ.ERO, ERO.NÄHDÄ, ERO.YMMÄRTÄÄ, ERO.HAVAITA, ERO.NUMEROT, ERO.ERI), *erota*-skeemoja 4 (EROTAA.IRTI, EROTA.KIRKKO, EROTA.PUOLISO, EROTA.ERI) ja *erottaa*-skeemoja 8 (EROTTA.LUOKITTELU, EROTTAA.HAVAITA2, EROTTAA.HAVAITA, EROTTAA.KONKRETIA, EROTTAA.ABSTRAKTI, EROTTAA.IRTI, EROTTAA.IRTI2 ja EROTTAA.ERI). Luvun 6.2.1 tarkastelussa samankaltaisuusmittarina toimii euklidinen etäisyys ja luvussa 6.2.2 canberra. Luvussa 6.2.3 kokoaan tutkimustulokset¹⁶³.

6.2.1 Klusterianalyysi A

Klusteristruktuurin – jossa yhdistämissääntönä on *Ward* ja samankaltaisuusmittarina *euklidinen etäisyys* – esittää kuvio 8. Kuvio 9 esittää klusteroinnin, jossa näkyvät Suzukin ja Shimodaira (2014) klusterointialgoritmin tuottamat todennäköisyysarvot p_{BP} ¹⁶⁴ (*bootstrap probability*) ja p_{AU} (*approximately unbiased*)¹⁶⁵. Ne on estimoitu 2000 toistolla eli *bootstrapilla*¹⁶⁶. Kutsun tätä klusterointia jatkossa klusteroinniksi A.

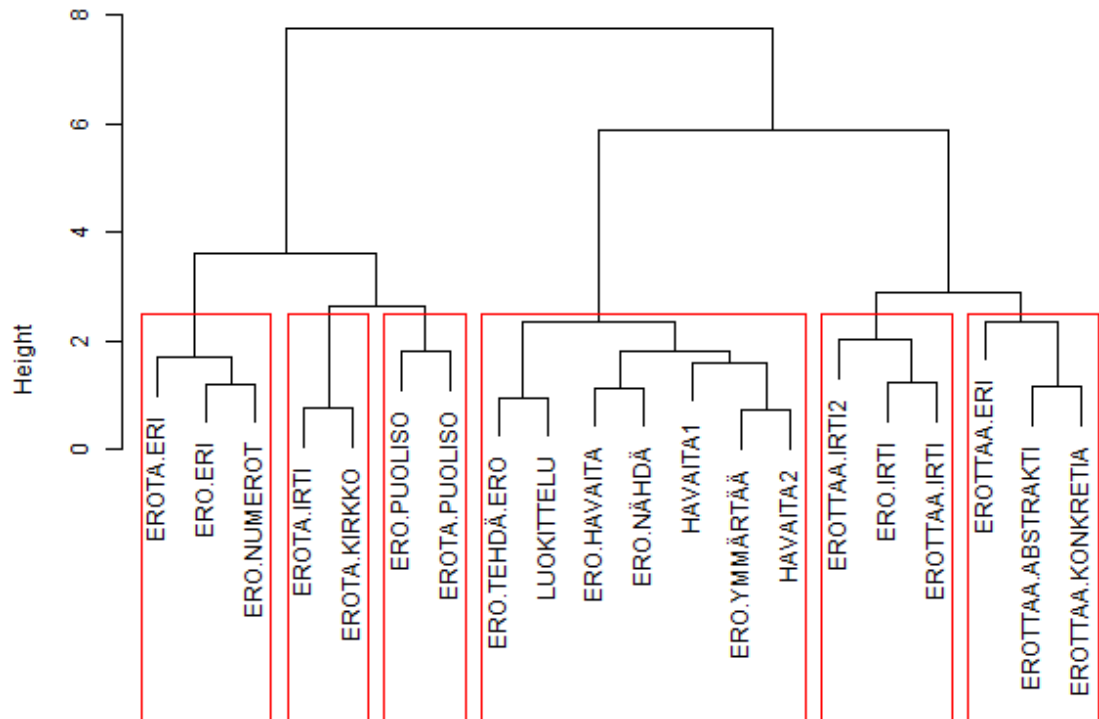
¹⁶² Käyttätymisprofiileihin syötetyt muuttujat luokkayhdistelyineen ovat **Y_SEM** (ELOTON, n=562, YKSILÖ, n=301; JOUKKO, n=248; KONKRETIA, n=159; INSTITUUTIO, n=118), **Z_SEM** (ELOTON, n=433, INSTITUUTIO, n=330; ELOLLINEN [**←** JOUKKO & YKSILÖ], n=140; KONKRETIA, n=134), **X_SEM** (ELOLLINEN [**←** JOUKKO & YKSILÖ], n=258; ELOTON, n=48; INSTITUUTIO, n=38), **Y_MUOTO** (NP, n=1023; NA [**←** NOLLA], n=275; PRON.RAK [**←** RAKENNE & PRON], n=237; PERS [**←** INKORPOROITU], n=126), **Z_MUOTO** (NP, n=928; NA, n=624; PRON.RAK [**←** RAKENNE & PRON], n=109), **X_MUOTO** (NA, n=947; NA.TRANS, n=280; NP [**←** RAKENNE & PRON], n=232; PERSON [**←** PERS, NOLLA & INKORPOROITU], n=202), **RESIPROOKKI** (LÄSNÄ, n=221; NA, n=1440), **PÄÄLUOKKA** (AKTIIVI, n=1498; PASSIIVI, n=163), **LUOKKA** (TRANSITIIVI, n=714; INTRANSITIIVI, n=609; KOPULA, n=338), **KIELTO** (NEUTRAALI, n=1264; KIELTEINEN, n=397), **MODUS** (INDIKATIIVI, n=1454; MUU, n=207), **TEMPUS** (PREESENS, n=1388; PREFEKT1, n=99; MUU, n=174), **VERBI_RAKENNE** (ERO.VERBI, n=1171; VERBIKETJU, n=369; MUU, n=121), **MODUS1** (KYKY, n=162; NA, n=1499), **MODUS2** (VÄLTÄMÄTTÖMYYS, n=175; MAHDOLLISUUS, n=136; NA, n=1350), **SY** (LÄSNÄ, n=93; NA, n=1568), **AIKA** (LÄSNÄ, n=191; NA, n=1470), **VASTAUS** (LÄSNÄ, n=174; NA, n=1487), **KVANTTORI** (LÄSNÄ, n=343; NA, n=1318).

¹⁶³ Klusterianalyysit on suoritettu *R*-kielen skriptillä (*Behavioral Profiles.R*, Gries 2010), jota olen muokannut tarpeisiini sopivaksi. Skripti suorittaa analyysit *R*-kielen komennoilla *hclust* ja *pvcust*. Näistä jälkimmäinen, joka on Suzukin ja Shimodairan (2014) *R*-paketti, tuottaa klustereiden todennäköisyysarvot.

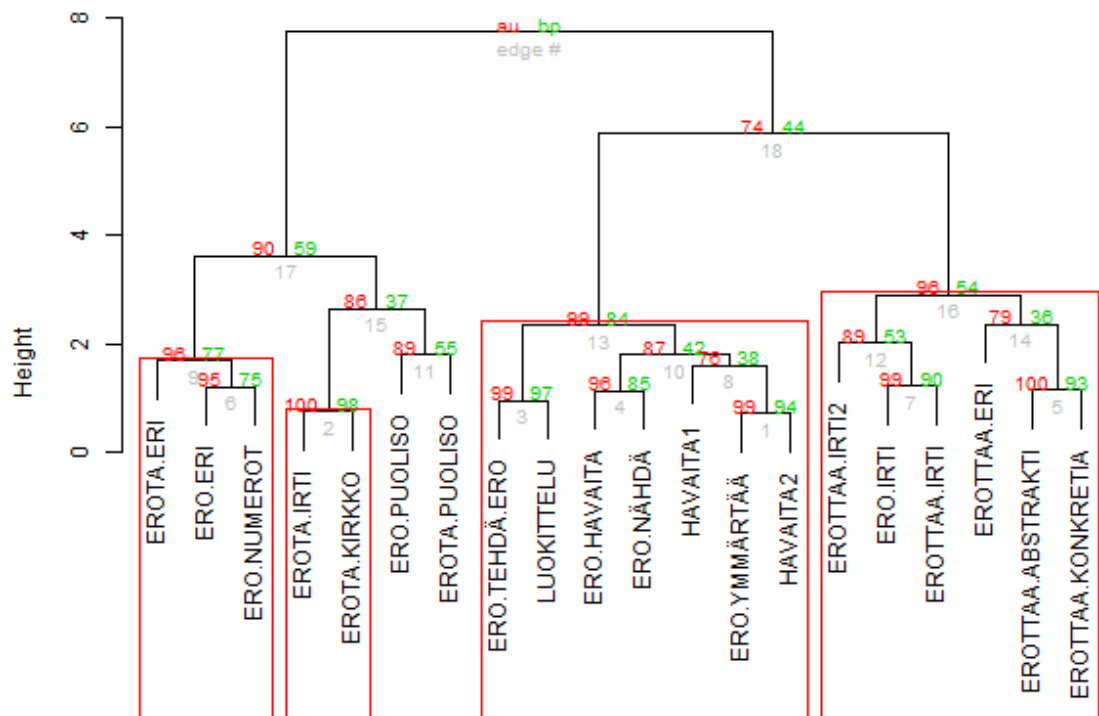
¹⁶⁴ Klusterikuvioden 9, 10, 11 ja 12 *p*-arvot ovat huonosti mustavalkotulosteesta erottuvilla vihreällä ja punaisella fontilla teknisistä syistä. En ole osannut vaihtaa niiden väriä.

¹⁶⁵ Suzuki ja Shimodaira (2014) kirjoittavat (kuvion 9) *p*-arvoista – joista vasemmanpuoleiset (punaiset) edustavat *AU*-arvoja ja oikeanpuoleiset (vihreät) *BP*-arvoja – seuraavasti: ”**AU** *p*-value [punainen, vasemmanpuoleinen arvo], which is computed by multiscale bootstrap resampling, is **a better** approximation to unbiased *p*-value than BP value computed by normal bootstrap resampling.” (Lihavointia ei ole alkuperäisessä lähteessä.)

¹⁶⁶ *Bootstrap*-viittaa tilastotieteessä käytettyyn algoritmiin, joka suorittaa otantajakaumaa. Yksinkertaistaen kyse on epävarmuuden tai todennäköisyyden estimoinnista keinotekoisella otannalla. Datasta (työssäni korpuseineistosta) otetaan *X* määrä (työssäni 2000) erilaista otantaa (*samples*), joiden pohjalta estimoidaan *p*-arvot. (ks. Baayen 2008: 146; yksityiskohtaisesti Arppe 2008: 135.) Huomautettakoon, että 2000 toistoa on suhteellisen vähän. Kirjallisuudessa tuloksia estimoidaan kymmenillä tuhansilla toistoilla.



Kuvio 8. Johdospesyeen klusterianalyysi A. Etäisyys: *euklidean*, yhdistämissääntö: *Ward*.



Kuvio 9. Klusterointi A jossa bootsratpillä määritellyt p_{AU} -arvot (vasemmanpuoleiset punaiset luvut) ja p_{BP} -arvot (oikeanpuoleiset vihreät luvut).

Kuviossa 8 (punaiset) laatikot signaloivat valitsemani klusteriluvun 6 tuottamaa klusterirakennetta. Valitsin 6 klusteria – joka antaa 80 % luokittelutarkkuuden PAM-testillä (ks. s.

159) – koska 6 klusterin rakenne tuntuu semanttisesti mielekkäältä ja koska haluan yksinkertaistaa analyysii¹⁶⁷. Keskimääräinen siluetin pituus (*average silhouette width*) on 0.269, joka Kaufmanin ja Rousseeuwin (1990: 88) ehdotuksen mukaan signaloii heikkoa tai keinotekoisia rakenteita¹⁶⁸. Ei ole oletettavaakaan, että syntaktisesti erilaisten sanojen klusterointi tuottaisi täydellisen homogeenisen rakenteen. Kuviossa 9 (punaiset) laatikot signaloivat klustereita, joiden *p*-arvo – olemassaolon todennäköisyys – ylittää tutkimukseni kriittisen merkitsevyystason $\alpha=0.05$ (jota kuvissa ”kääntäen” vastaa 95 %). Klusteroitavana korpusaineisto tukee erityisesti näiden klusterien olemassaoloa. Tarkastelen pelkästään p_{AU} -arvoja (vasemmanpuoleiset punaiset luvut), jotka Suzukin ja Shimodairan (2014) mukaan ovat p_{BP} -arvoja luotettavimpia (ks. s. 161 alav. 163). Klusterianalyysissä ei yleensä tehdä pitkälle meneviä merkittävyysoletuksia, minkä takia suhtaudun *p*-arvoihin – ja lisäksi klusterianalyysin tuloksiin – lähtökohtaisesti suuntaa-antavina tuloksina.

Viittaan jatkossa kuvioiden 8 ja 9 esittämään klusterointirakenteeseen notaatiolla A-klusterointi. Kuviossa 8 esitellyt 6 klusteria, joiden notaatioissa hakasulut signaloivat yksittäistä klusteria ja &-merkki kahden skeeman yhdistymistä – ovat:

Klusteri A1: [[ERO.NUMEROT & ERO.ERI], EROTA.ERI]

Klusteri A2: [EROTA.IRTI & EROTA.KIRKKO]

Klusteri A3: [ERO.PUOLISO & EROTA.PUOLISO]

Klusteri A4: [[[[ERO.YMMÄRTÄÄ & EROTTAA.HAVAITA2], EROTTAA.HAVAITA1],
ERO.HAVAITA & ERO.NÄHDÄ], [ERO.TEHDÄ.ERO &
EROTTA.LUOKITTELU]]

Klusteri A5: [[ERO.IRTI & EROTTAA.IRTI], EROTTAA.IRTI2]

Klusteri A6: [[EROTTA.ABSTRAKTI & EROTTAA.KONKRETIA], EROTTAA.ERI]

Klustereista voidaan laskea monenlaisia tunnuslukuja, esimerkiksi *z*-arvoja (*z-score* jolla tarkastellaan yksittäisen klusterin eri jäsenten samankaltaisuutta) ja *f*-arvoja (*f-value* jolla määritellään klustereiden homogeenisuutta) (Divjak & Gries 2006, 2009b; Gries 2012). Tässä työssä tarkastelen klustereiden välisiä suhteita *t*-arvoilla (ks. Divjak & Gries 2008:

¹⁶⁷ 7 klusterin rakenne, jossa [EROTTA.ERI] muodostaa oman klusterin, antaa PAM-tarkkuudeksi 80 % ja 8 klusterin rakenne, jossa klusterista A4 irtautuu klusteri A4a [ERO.TEHDÄ.ERO ja EROTTAA.LUOKITTELU] antaa PAM-tarkkuudeksi 35 %.

¹⁶⁸ Kaufmanin ja Rousseeuwin (1990: 88) sanalliset ehdotukset kynnysarvoille ovat: 0.71–1.0 ‘a strong structure has been found’, 0.51–0.70 ‘a reasonable structure has been found’, 0.26–0.50 ‘the structure is weak and could be artificial’ ja < 0.25 ‘no substantial structure has been found’. Vaikka esimerkiksi Gries käyttää Kaufmanin ja Rousseeuwin (1990) lähteenään usein, en ole kielitieteellisissä tutkimuksissa tavannut kynnysarvojen sanallista soveltamista, enkä siksi osaa arvioida niiden paikkansa pitävyyttä johdospeseyeen polysemiaa käsittelevään semanttiseen aineistoon (vrt. kuitenkin Gries [2012: 66], joka kirjoittaa: ”In order to determine the most likely number of groups the synonymous verbs fall into, one can use the measure of average silhouette widths. Silhouette widths are a statistic that compare within and between cluster similarities, and the higher an average silhouette width for a particular cluster solution, the better that cluster solution.”)

192–193, 203). *t*-arvojen avulla voidaan selvittää, mitkä ID-merkit kytkeytyvät tiettyyn klusteriin eli mitkä alakategoriat ovat aliedustettuina missäkin klusterissa (suuri negatiivinen arvo) ja mitkä yliedustettuina (suuri positiivinen arvo) (Gries 2012: 67)¹⁶⁹. *t*-arvot, jotka ovat kulloinkin tarkastelussaan olevassa klusterissa suuria (positiivinen arvo) ja muissa klustereissa pieniä (negatiivinen arvo), on lihavoitu. *t*-arvojen negatiivinen huippu klusteroinnissa A on (-1.79) ja positiivinen huippu (+2.25)¹⁷⁰.

Klusteriin A1 kuuluvat skeemat EROTA.ERI sekä aikaisemmin klusteroituvat (eli suhteellisesti enemmän samankaltaiset) skeemat ERO.NUMEROT ja ERO.ERI. Semanttisesti klusteria voisi luonnehtia (luvuissa 1.3 ja 3.3.1 esitellyn) 'ominaisuuden tai laadun eroamisen' pesyeskeeman toteutumaksi. Muihin klustereihin verrattuna klusterin A1 skeemoille on ominaista esiintyä kopulalauseessa (LUOKKA_KOPULA = +1.66) ilman modaalista apuverbiä (VERBI_RAKENNE_ERO.VERBI = +1.66) ja kausatiivista agenttia (X_MUOTO_NA = +1.34) siten, että lause on aktiivissa (PÄÄLUOKKA_AKTIIVI = + 0.82) ja aikamuotona on preesens (TEMPUS_PREESENS = +0.93). Y- ja Z-argumentteina esiintyy erityisesti elottomia tarkoitteita (Y_SEM_ELOTON = +1.05; Z_SEM_ELOTON = +1.03). Klusterin korkein *t*-arvo on luokalla KVANTTORI_LÄSNÄ (+1.79), jota havainnollistaa lause *Yhteisöllisyydellä_{Y_SEM_ELOTON} ja anarkialla_{Z_SEM_ELOTON} on vissi ero* (K1631). Tässä on läsnä useita klusterille A1 tyypillisiä ominaisuuksia. Klusterin vahvimpia negatiivisia *t*-arvoja ovat MODUS2_MAHDOLLISUUS (-0.86) – joka signaloi sitä, ettei skeeman toteutumisissa tapaa modaalista mahdollisuutta (tai mahdolltomuutta) ilmoittavia apuverbejä – sekä AIKA_LÄSNÄ (-0.91), sillä ominaisuuden ja laadun eroavaisuutta predikoivien lauseiden merkitystausta ei yleensä sisällä temporaalista hahmoa. Klusterin A1 olemassaolo on todennäköistä ($p_{AU} = 96\%$) merkitsevyystasolla $\alpha=0.05$.

Klusteri A2 [EROTA.IRTI, EROTA.KIRKKO] yhdistyy ensimmäisten joukossa, mikä on vihje siitä, että klusteri saattaa olla suhteellisen homogeeninen. Klusterin skeemat EROTA.IRTI ja EROTA.KIRKKO esiintyvät intransitiivilauseessa (+1.91; X_MUOTO_NA = +1.46) etenkin imperfektissä tai pluskvamperfektissä (TEMPUS_MUU = +1.01) ja suosivat rektioadverbiaalissa luokkaa Z_SEM_INSTITUUTIO (+1.73). Niiden toteumat hakeutuvat verbiketjuihin (VERBI_RAKENNE_VERBIKETJU = +0.88), jotka ilmaisevat mahdollisuuteen liittyvää modaalisuutta (MODUS2_MAHDOLLISUUS = +1.68). Ilmipannut agentit ovat yksilötarkoitteisia

¹⁶⁹ Korkean *t*-arvon kytkeytyminen tiettyyn klusteriin signaloi siis, että tähän klusteriin kuuluvien skeemojen toteutumisissa esiintyy verrattain usein nimenomaisen alakategorian toteutumia. *t*-arvot olen laskenut kirjoittamallani skriptillä *t-arvot.R*, jonka kaavan klusterianalyysiin esittävät Backhaus, Erichson, Plinke ja Weiber (2003: 310–312; ks. myös Gries 2012: 67). Kaava on: (ID-merkin *x* keskiarvo klusterissa *c* – ID-merkin *x* keskiarvo) / ID-merkin *x* keskihajonta.

¹⁷⁰ *t*-arvot ovat täysin riippuvaisia kulloisestakin klusterointitarkastuksesta, eivätkä kahden eri klusteroinnin *t*-arvot ole keskenään vertailukelpoisia.

olioita (Y_SEM_YKSIÖ = +1.74), joilla on taipumus inkorporoitua verbiin tai esiintyä persoonapronominissa (Y_MUOTO_PERS = +1.88). Lauseissa on usein aikaa spesifioivia adverbiaaleja (AIKA_LÄSNÄ = +1.37). Moni näistä ominaisuuksista on läsnä lauseessa *Lapsi Y_SEM_YKSIÖ voi erota VERBI_RAKENNE VERBIKETJU kirkosta Z_SEM_INSTITUUTIO aikuisena AIKA, jos haluaa* (K577). Klusterin negatiivisimpien *t*-arvojen joukossa ovat Y_SEM_ELTON (-1.09) ja Z_SEM_ELTON (-1.01), mikä korostaa sitä, että klusterissa on kyse ihmisille tapahtuvasta toiminnasta. Klusterin olemassaolo on oletettavaa ($p_{AU} = 100\%$). Klusterin olemassaolo vahvistaa luvussa 5.2 tehtyjä havaintoja.

Klusterin A3 skeemat ERO.PUOLISO ja EROTA.PUOLISO karttavat kaikkia ilmipantuja argumenttirooleja (Y_MUOTO_NA = +1.83, Z_MUOTO_NA = +1.80; X_MUOTO_NA = +1.10) ja hakeutuvat intransitiivilauseisiin (LUOKKA_INTRANSITIIVI = +1.02) sekä perfektiin (TEMPUS_PERFEKTI = +1.04). Ilmipannun agentin spesifioi eritoten joukkotarkoitteinen olento (Y_SEM_JOUKKO = +2.35), lähteen taas yksilö tai joukko (Z_SEM_ELOLLINEN = +1.24). Esiintymissä on usein toiminnan syyhyn liittyviä adverbiaaleja (SYYS_LÄSNÄ = +1.01). Monet mainituista ominaisuuksista esiintyvät lauseessa *Lesbopareista on eronnut-TEMPUS_PERFEKTI 24 prosenttia Y_SEM_JOUKKO* (K972). Klusteria voi pitää pesyeskeema 'parisuh-teesta eroaminen' toteutumana. Klusteri ei ole olemassa ($p_{AU}=89\%$) merkitsevyystasolla $\alpha=0.05$. Klusteri yhdistyykin kuvioissa 8 ja 9 suhteellisen myöhään, mikä ehkä signaloi verbin ja nominin perustavanlaatuista eroa kielenyksikkönä.

Klusteri A4 koostuu useasta pienemmästä klusterista, joista merkitsevyystasolla $\alpha=0.05$ olemassa ovat klusteri A4a [ERO.TEHDÄ.ERO ja EROTTAA.LUOKITTELU, $p_{AU}=99\%$], jonka skeemoja ilmentävät lauseet *On tärkeää tehdä ero vapaaehtoisen ja pakotetun seksityön välille* (K1615, *ero*) ja *Tässäkin keskustelussa joillekin on vaikea erottaa yliopistoja, lukioita ja ylä-asteita toisistaan* (K293, *erottaa*), klusteri A4b [ERO.HAVAITA ja ERO.NÄHDÄ, $p_{AU}=96\%$] ja klusteri A4c [ERO.YMMÄRTÄÄ ja EROTTAA.HAVAITA2, $p_{AU}=99\%$], jonka skeemoja ilmentävät esimerkit *On paljon haitallisempaa jos sairaanhoitaja ei ymmärrä eroa mikrogramman ja milligramman välillä* (K1082, *ero*) ja *Oliko niin, ettet osaa erottaa edes suomenruotsaista ruotsinsuomalaisesta* (K313, *erottaa*). Skeema EROTTAA.HAVAITA1 yhdistyy klusteriin [ERO.YMMÄRTÄÄ ja EROTTAA.HAVAITA2] näitä ylempänä, mikä heijastelee semanttista kaltaisuutta mutta myös välimatkaa. Havainto tukee yksimuuttujaisen analyysin tuloksia (ks. luku 5.3). Klusteri [[ERO.YMMÄRTÄÄ ja EROTTAA.HAVAITA2], EROTTAA.HAVAITA1] ei olekaan kriittisyystasolla $\alpha=0.05$ tilastollisesti merkitsevä, mutta klusteri A4 kokonaisuudessaan on ($p_{AU}=99\%$).

Klusterin A4 korpusesimerkit esiintyvät etenkin kielteisissä (KIELTO_LÄSNÄ = +1.09)

transitiivilauseissa (LUOKKA_TRANSITIIVI = +0.73), joissa ihmistarkoitteinen agentti (X_SEM_ELOLLINEN = +1.02) esiintyy ilmipantuna persoonapronominissa tai verbiin inkorporoituneena tai nollapersoonassa (X_MUOTO_PERSON = +1.26). Muita klustereita useammin klusterin skeemojen toteumissa on läsnä kykyä ilmoittava modaalinen apuverbi (MODUS1_KYKY_LÄSNÄ = +1.12). Y-argumentin – eli patientin – spesifioi usein eloton entiteetti (Y_SEM_ELOTON = +0.82), esimerkiksi ajatus tai käsite. Etenkin klusterissa A4c [ERO.YMMÄRTÄÄ ja EROTTAA.HAVAITA2] esiintyy patientteina PRON ja RAKENNE luokkien toteutumia eli *tämä*-, *se*- tai *joka*-edussanaisia lausekkeita tai lausetäydennyksiä (Y_MUOTO_PRON.RAK = +0.95), jotka kytkeytyvät resiprookkisella adverbialilla (RESIPROOKKI_LÄSNÄ = +0.71) lähteeseen. Se on usein konkreettinen tai abstrakti entiteetti (Z_SEM_KONKRETTIA = +0.55; Z_SEM_ELOTON = +0.67). Klusteria A4 voi kokonaisuudessaan luonnehtia (ks. luvut 1.3 ja 3.3.3) ’kognitiivisen toiminnan ja havainnon skeemaksi’. Se, että klusterianalyysi on osannut ryhmitellä nominin ja verbin skeemat klusteriin A4, johtuu sekä tavasta, jolla olen koodannut aineistoni, että myös siitä, että näillä ilmaisutyypeillä todella ehkä on kaltaisuutta käyttäytymisprofiilien tasolla. Kyse ei kuitenkaan ole *ero*-nominin ja *erottaa*-verbin suorasta assosiaatiosta, vaan skeemojen assosiaatiosta. Semanttinen kaltaisuus syntyy esimerkiksi skeeman EROTTAA.HAVAITA2 ko(n)tekstuaalisen profiilin assosioitumista sellaisten transitiiviverbien ko(n)tekstuaalisiin profileihin, joiden vaikutusalassa *ero*-lausekkeet ovat. Kyse on siis ilmaisutyyppien tai konstruktioiden (vrt. Goldberg 1995) välisestä semanttis-syntaktisesta kaltaisuudesta, jota käyttäytymisprofiilit heijastavat, kuten minimiparissa *enää ei ymmärretä eroa karhun ja ketun jätösten välillä* (ERO.YMMÄRTÄÄ) : *enää ei osata erottaa karhun ja ketun jätöksiä toisistaan* (EROTTA.HAVAITA2). Kaltaisuus ei tarkoita synonymiaa. *ero*-nominin skeemat ERO.HAVAITA, ERO.NÄHDÄ, ERO.TEHDÄ.ERO ja ERO.YMMÄRTÄÄ ilmaisevat pikemminkin mentaalista prosessia, kun taas *erottaa*-skeemat usein lisäksi aistihavaintoa. Verbit *erottaa* ja *ymmärtää* tai *maistaa* voivat olla korvattavissa joissakin konteksteissa mutta synonyymejä ne tuskin ovat.

Klusteriin A5 ensin yhdistyvä klusteri A5a [ERO.IRTI ja EROTTAA.IRTI] on olemassa ($p_{AU}=99\%$) kriittisyystasolla $\alpha=0.05$. Tämä kytkös on kiinnostava, sillä se on tutkimushypoteesini mukainen. Aikainen yhdistyminen vihjaa, että skeemojen samankaltaisuus on suurempaa kuin ylemmän tason klusterissa A5b [[ERO.IRTI & EROTTAA.IRTI], EROTTAA.IRTI2], jonka olemassaolo ei olekaan todennäköistä kriittisyystasolla $\alpha=0.05$. Myös luvussa 5.3.6 suoritetussa klusterianalyysissä skeemat EROTTAA.IRTI ja EROTTAA.IRTI2 ryhmittyvät yhteen,

tosin viimeisenä, mikä vihjailee paitsi semanttista kaltaisuutta myös klusterin sisäistä heterogeenisuutta, jota työni laajuuden vuoksi en erikseen estimoi. Ylemmän tason klusteri [klusteri A5 + klusteri A6] on jälleen merkitsevä ($p_{AU}=96\%$).

Klusterin A5 korpustoteumat esiintyvät usein agentittomassa (X_MUOTO_NA.TRANS = +1.69) passiivilauseissa (PÄÄLUOKKA_PASSIIVI = +1.37), joissa on esillä välttämättömyyttä ilmaiseva modaalinen apuverbi (MODUS2_NECESSITY = **+1.41**). Myös syytä spesifioiva adverbialli (SYYS_LÄSNÄ = +1.40) on usein läsnä. Preesens ei ole klusterin ainoa aikamuoto (TEMPUS_PERFEKTI = +1.12; TEMPUS_MUU = +0.77). Kaikissa argumenttirooleissa esiintyy luokkaan INSTITUUTIO koodattuja tarkoituksia (Z_SEM_INSTITUUTIO = +1.64, Y_SEM_INSTITUUTIO = **+1.46** ja X_SEM_INSTITUUTIO = +0.72) – mikä heijastaa etenkin skeeman EROTTA.ABSTRAKTI vaikutusta (ks. luku 5.3). Patientteina esiintyy suhteellisen usein luokkaan NP (Y_MUOTO_NP = +1.0) koodattuja yksilötarkoituksia (Y_SEM_YKSIÖ = +0.95). Ominaisuus kytkee klusteria A5 klusteriin A2 [EROTTA.ABSTRAKTI ja EROTTA.KONKRETIA; Y_SEM_YKSIÖ = +1.74], samoin kuin aikaa ilmoittavan adverbialin läsnäolo (klusteri A5, AIKA_LÄSNÄ = +1.00; klusteri A2, AIKA_LÄSNÄ = +1.37). Monet mainituista ominaisuuksista ovat läsnä esimerkissä *Jos vanhemmat kieltäytyvät tulemasta, silloin häirikkö* Y_MUOTO_NP~Y_SEM_YKSIÖ **erote-**
taan PÄÄLUOKKA_PASSIIVI *koulusta* Z_SEM_INSTITUUTIO (K103).

Klusteri A6 [[EROTTA.ABSTRAKTI ja EROTTA.KONKRETIA], EROTTA.ERI] olemassaolon todennäköisyys on epävarmaa ($p_{AU}=79\%$). Siihen ensin yhdistyvän klusterin A6a [EROTTA.ABSTRAKTI ja EROTTA.KONKRETIA] olemassaolon todennäköisyys sen sijaan on merkitsevää ($p_{AU}=100\%$). Tulos tukee yksimuuttujaisen analyysin tuloksia ja luvussa 5.3.6 suoritettua klusterointia. Klusteri A6 sisältääkin heterogeenisiä aineksia, sillä vaikka klusteri A6a [EROTTA.ABSTRAKTI ja EROTTA.KONKRETIA] on yksimuuttujaisen analyysin perusteella (ks. luku 5.4.4) suhteellisen homogeeninen, skeema EROTTA.ERI kasvattaa klusterin ”hajontaa”. Yksi- ja monimuuttujaisen analyysin tulokset eivät kuitenkaan ole suoraan vertailukelpoisia, sillä klusteroinnin pohjalla oleviin käyttäytymisprofileihin on sisällytetty huomattavasti enemmän luokkia. Vaikka muuttujia on enemmän, klusteroinnin tulokset ovat kuitenkin kiinnostavan samanlaisia. Tämä vihjaa siihen, että skeemojen vaihtelun ydinmekaniikasta iso osa on selitettävissä jo X-, Y-, Z-argumenttien nojalla.

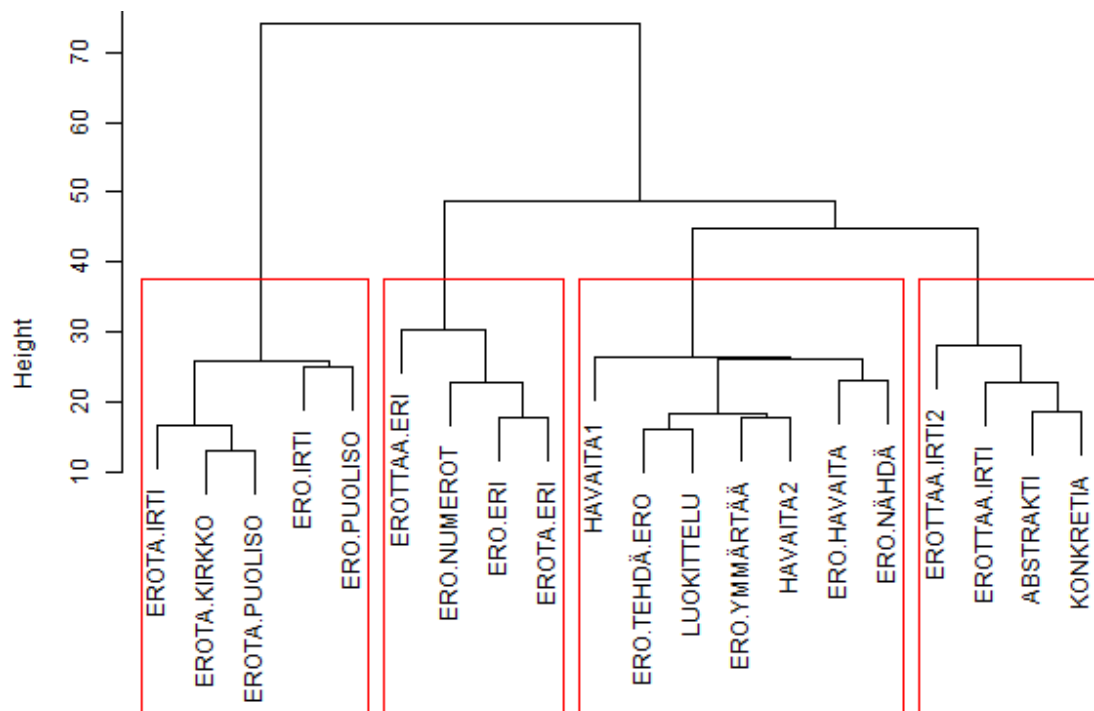
Klusterin A6 esiintymissä – eli skeemassa EROTTA.ERI – on usein eloton agentti (X_SEM_ELTON = **+1.97**), joka koodautuu luokkana NP (X_MUOTO_NP). Luokka NP sisältää yksimuuttujaisessa analyysissä skeemaan EROTTA.ERI kytkeytyneet, subjektin paikalla sijaitsevat luokat PRON (pronominit) ja RAKENNE (*se*-tukipronomini + *että*-lause -rakenne). Jälkimmäinen on usein luokan VASTAUS_LÄSNÄ (+1.03) toteuma, kuten esimerkissä *Työn*

*erottaa harrastuksesta se, että siitä saa palkkaa*_{X_SEM_ELOTON} ~ _{X_MUOTO_RAKENNE} ~ _{VASTAUS_LÄSNÄ}. (K275). Tämä ominaisuus kytkee klusteria A6 – tai skeemaa EROTTAA.ERI – klusteriin A1 [[ERO.NUMEROT & ERO.ERI], EROTA.ERI], jossa on luokan VASTAUS_LÄSNÄ toiseksi suurin *t*-arvo, +0.42, muiden klustereiden sisältäessä negatiivisen *t*-arvon. Passiivin painotus (PÄÄLUOKKA_PASSIIVI = +0.78) yhdistää klusteria A6 – tai skeemoja EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA – klusteriin A5 [ERO.IRTI ja EROTTAA.IRTI] (PÄÄLUOKKA_PASSIIVI = +1.37) vastakohtana klusterille A1 [[ERO.NUMEROT & ERO.ERI], EROTA.ERI] (PÄÄLUOKKA_PASSIIVI = -0.82). Patientin spesifioi usein konkreettinen entiteetti (Y_SEM_KONKRETIA = +1.00), mikä heijastaa etenkin skeeman EROTTAA.KONKRETIA vaikutusta.

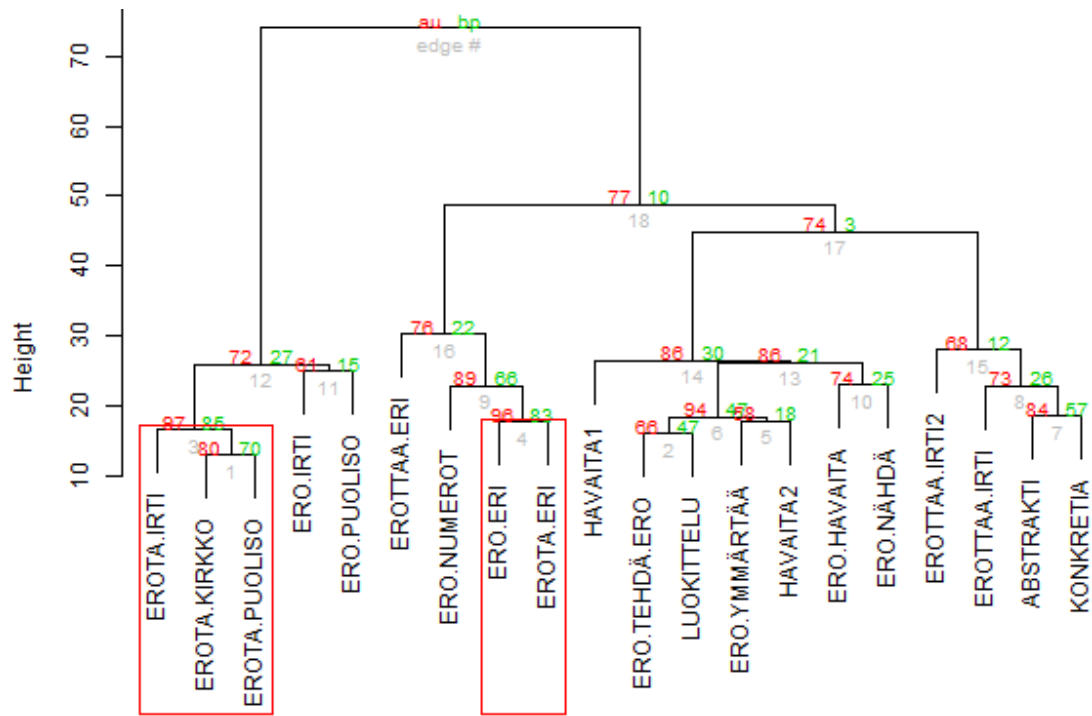
En analysoi klusterointia A tätä enempää, vaan siirryn tarkastelemaan klusterointia B. Palaan klusteroinnin A tuloksiin yhteenvedon yhteydessä luvussa 6.2.3.

6.2.2 Klusterianalyysi B

Tämän luvun ja edellisen luvun klusterointiasetukset ovat identtiset samankaltaisuusmittaria lukuun ottamatta. Klusteristruktuurin, jossa yhdistämissääntönä on *Ward* ja samankaltaisuusmittarina *canberra*, esittää kuvio 10. Kuvio 11 esittää klusteroinnin, josta ilmenevät todennäköisyysarvot p_{BP} ja p_{AU} 2000 toistolla eli *bootstrapilla*. Näistä tarkastelen pelkästään p_{BP} -arvoja eli vasemmanpuoleisia lukuja. Kutsun tätä klusterointia klusterointi B:ksi.



Kuvio 10. Johdospesyeen klusterianalyysi B. Etäisyys: *canberra*, yhdistämissääntö: *Ward*.



Kuvio 11. Klusterointi B p_{AU} -arvoilla (vasemmanpuoleiset punaiset luvut) ja p_{BP} -arvoilla (oikeanpuoleiset vihreät luvut).¹⁷¹

Kuten klusterikuvioista 10 näkyy, valitsin klusteriluvuksi 4, joka tuottaa semanttisesti suhteellisen mielekkään struktuurin ja antaa 80 % luokittelutarkkuuden PAM-testillä. Keskimääräinen siluetin pituus (*average silhouette width*) on 0.15685, joka Kaufmanin ja Rousseeuwin (1990: 88) mukaan signaloi, ettei huomattavaa rakennetta ole löytynyt¹⁷². Kielitieteellisesti kuvion 10 struktuuri tuntuu kuitenkin eräiltä osin kuvio 8 mielekkäämmältä. Kuvion 10 klusterit vasemmalta oikealle (ja alhaalta ylös) luettuna ovat:

Klusteri B1: [[[EROTA.KIRKKO & EROTA.PUOLISO], EROTA.IRT1], [ERO.IRT1 & ERO.PUOLISO]]

Klusteri B2: [[[EROTA.ERI & ERO.ERI], ERO.NUMEROT], EROTTAA.ERI]

Klusteri B3: [[[ERO.TEHDÄ.ERO & EROTTAA.LUOKITTELU], [ERO.YMMÄRTÄÄ & HAVAITA2]], [ERO.HAVAITA & ERO.NÄHDÄ], HAVAITA1]

Klusteri B4: [[[EROTTA.ABSTRAKTI & EROTTAA.KONKRETIA], EROTTAA.IRT1], EROTTAA.IRT2]

Klusterointiratkaisuissa A ja B on joitakin eroja. Nyt skeema EROTTAA.ERI yhdistyy klusteroinin A klusteriin A1 [[ERO.NUMEROT ja ERO.ERI], EROTA.ERI], jota klusteri- ja struktuurissa B vastaa klusteri B2 [[[EROTA.ERI ja ERO.ERI], ERO.NUMEROT], EROTTAA.ERI]. Lisäksi skeema ERO.ERI ei yhdisty ensimmäisenä skeemaan ERO.PUOLISO, vaan skeemaan EROTA.ERI muodostaen klusterin B2a [EROTA.ERI ja ERO.ERI]. Tämä vihjaa,

¹⁷¹ Heikosti mustavalkotulosteissa näkyvä vihreä ja punainen fontinväri johtuu teknisistä syistä (ongelmista).

¹⁷² Siluettien pituudet aineistooni sovelletuissa canberra-testeissä ovat kautta linjan selvästi pienempiä.

että nämä skeemat ovat klusterin B2 skeemoista kaikista samankaltaisimpia. Klusteri B2a olemassaolo onkin merkitsevä ($p_{BP}=96\%$).

Klusteri B3 organisoituu hieman eri tavalla kuin A-klusteroinnin vastaava klusteri A4, joka ilmoittaa kognitiivista prosessia tai aistimista. ”Sisäklusterit” ovat kuitenkin samat, järjestys vain eri. Klusterissa B1 [[[EROTA.KIRKKO ja EROTA.PUOLISO], EROTA.IRTI], [ERO.IRTI ja ERO.PUOLISO]] – jonka skeemat ilmentävät monenlaisten suhteiden päättymistä – yhdistyvät ensimmäisenä skeemat EROTA.PUOLISO ja EROTA.KIRKKO. Niihin yhdistyy ylempänä skeema EROTA.IRTI, kunnes näiden muostama intransitiivinen klusteri yhdistyy kopulaklusteriin [ERO.IRTI ja ERO.PUOLISO]. Kopulaklusterin olemassaolon todennäköisyys ($p_{BP}=61\%$) on heikkoa, mutta intransitiiviklusterin merkitsevää ($p_{BP}=97\%$). Skeema ERO.IRTI ei siis enää linkity skeemaan EROTTAA.IRTI, vaan se muodostaa klusterin B1, johon kuuluvat klusteroinnin A klusterit A2 [EROTA.IRTI ja EROTA.KIRKKO] ja A3 [ERO.PUOLISO ja EROTA.PUOLISO]. Lisäksi klusterointi B poikkeaa A:sta siinä, että klusterit A5 ja A6 – skeemoja EROTTAA.IRTI ja ERO.ERI lukuun ottamatta – fuusioituvat B:ssä klusteriksi B4 [[EROTTAABSTRAKTI ja EROTTAAKONKRETIA], EROTTAA.IRTI], EROTTAA.IRTI2]. Tämä vastaa yksimuuttujaisen analyysin (ks. luku 5.3.5) passiivimyönteistä ryhmää. Yhdistymisjärjestys vihjaa, että klusteri B4a [EROTTAABSTRAKTI ja EROTTAAKONKRETIA] on näistä homogeenisin. Struktuurissa B heijastuu A:ta vahvemmin se, koodaavatko skeemojen toteumat kopula- tai intransitiivilauseetta (klusterit B1 ja B2) vai transitiivilauseetta (klusterit B3 ja B4).

Klusteroinnin t -arvot liikkuvat tällä kertaa välillä (-1.75) ja $(+1.75)$. Aloitan vasemmanpuoleisesta klusterista B1, johon kuuluvat parisuhteesta (ERO.PUOLISO, EROTA.PUOLISO), työsuhteesta tai vastaavasta (ERO.IRTI, EROTA.IRTI) ja kirkosta (EROTA.KIRKKO) eroamisen luentoja spesifioivat skeemat.

Klusterin skeemojen korpustoteutumista valtaosa esiintyy intransitiivilauseessa (LUOKKA_INTRANSITIIVI = **+1.11**), menneessä aikamuodossa (TEMPUS_MUU = **+1.06**; TEMPUS_PREESENS = -1.01), siten että agentin (Y-argumentti) spesifioi yksilö- tai joukkotarkoittinen olio (Y_SEM_YKSILÖ = **+1.23**; Y_SEM_JOUKKO = $+0.97$; Y_SEM_ELOTON = -1.09). Aikaan ja syyhyn liittyvät adverbialit ovat kaikista yleisimpiä nimenomaan klusterin B1 skeemoissa (SYY_LÄSNÄ = $+1.17$; AIKA_LÄSNÄ = $+0.86$). Lähde, milloin ilmipantuna, edustaa tyypillisesti luokkaa instituutio (Z_SEM_INSITUUTIO = $+0.69$). Klusteriin kytkeytyvät kaikkien argumenttiroolien ilmipanemattomuus, joka heijastaa PUOLISO-skeemojen käyttäytymistä (X_MUOTO_NA = $+0.97$, Y_MUOTO_NA = **+0.95** ja Z_MUOTO_NA = **+0.73**). Negatiivisista t -arvoista mainittakoon klustereista suurin resiprookin (*toisistaan ~ välillä*) vastustus (RESIPROOKKI_LÄSNÄ = -0.8). t -arvojen valossa klusterissa

yhdistyy edellisen klusteroinnin klusterien A2 [EROTA.IRTI, EROTA.KIRKKO] ja A3 [ERO.PUOLISO, EROTA.PUOLISO] mittaustulokset.

Myös klusterin B2 [[[EROTA.ERI & ERO.ERI], ERO.NUMEROT], EROTTAA.ERI] – joka heijastaa ’laadun ja ominaisuuden eroavaisuuden’ pesyeskeemaa – *t*-arvojen kärki muistuttaa struktuurin A klusterin A1 [ERO.NUMEROT & ERO.ERI], EROTA.ERI] mittaustuloksia. Tulokset ovat eräitten luokkien kohdalla sikäli homogeenisemmat, että skeema EROTTAA.ERI sijoittuu klusteriin B2, joskin sen läsnäolo tuo myös hajontaa transitiiviominaisuuksien kautta. Klusterin korpusesiintymät hakeutuvat preesenssissä olevaan (TEMPUS_PREESENS = +0.99) yksiverbiseen (VERBI_RAKENNE_ERO.VERBI = +1.18) kopulauseeseen (LUOKKA_KOPULA = +1.14), jossa usein on läsnä ominaisuuden eroavuuden intensiteettiä määrittävä adjektiivi tai adverbi KVANTTORI_LÄSNÄ (+1.30) ja / tai luokkaan VASTAUS koodattu (VASTAUS_LÄSNÄ = +1.29) täydennys, joka ilmoittaa entiteettien eroavaisuuden proposition (ks. luku 4.4.3). Klusteri B2 karttaa modaalisia aineksia (MODUS1_KYKY = -0.69; MODUS2_VÄLTÄMÄTTÖMYYS = -0.72 ja MODUS2_MAHdollisuus = -0.72), perfektii ja menneitä aikamuotoja (TEMPUS_PERFEKTI = -0.84; TEMPUS_MUU = -0.8), esiintymästä muussa kuin indikatiivissa (MODUS_MUU = -0.83) sekä esiintymistä passiivissa (PÄÄLUOKKA_PASSIIVI = -0.82).

Havaitsemista ja kognitiivista toimintaa edustavista skeemoista koostuvan klusterin B3 tulokset vastaavat pitkälti A-klusteroinnin klusterin A4 *t*-arvoja, minkä takia en käy niitä toistamaan. Klusterin B4 [[[EROTTAABSTRAKTI & EROTTAA.KONKRETIA], EROTTAA.IRTI], EROTTAA.IRTI2] *t*-arvot taas vastaavat A-klusteroinnin klustereiden A5 ja A6 tuloksia, sillä erotuksella, että klusteria B4 ei sotke skeeman EROTTAA.ERI poikkeava käyttäytymisprofiili.

6.2.3 Yhteenveto klusterianalyysistä

Työssäni sovellettu käyttäytymisprofiilipohjainen klusterianalyysi on menetelmä, joka tuottaa klusterirakenteen ilman inhimillistä maalaisjärkeä, yksinomaan valittujen algoritmien ja analyysiin syötettyjen käyttäytymisprofiilien varassa. Siksi klusterirakenteiden skeemaryhmittelyt ovat yllättävänkin positiivisia. Ne paljastavat, että *ero*-lekseemien ”eri merkityksiä” projisoivien skeemojen välillä on johdospesyeen tasolla kaltaisuutta, joka pätee kuitenkin vain koodaussysteemiä läpi tarkasteltuun HS.fi-aineistoon. Kaltaisuus nousee käyttäytymisprofiilien analogiasta ja ilmenee klusterianalyysin tuottamina, lingvistisesti kohtalaisen mielekkäinä ryhmittelyinä. Koska luvuissa 6.2.2 ja 6.2.3 klusteroidut käyttäytymisprofiilit sisältävät 53 ID-merkkiä 19 muuttujatyypistä, tulokset antavat huomattavasti monipuolisemman kuvan *ero*-pesyeen skeemoista kuin yksimuuttujainen analyysi. Analyysivaiheet eivät

kuitenkaan ole suoraan vertailukelpoisia, sillä sekä analyysityökalut että muuttujat ovat eri vaiheissa erilaisia. Potilas on kuitenkin sama ja lisäksi tulokset sen verran samansuuntaisia, että analyysivaiheiden vertailu tuntuu relevantilta. Teen seuraavaksi yhteenvetoa löydöistä.

Sekä A että etenkin B-klusterointi onnistuu poimimaan klusterin B2, jota voi nimittää 'ominaisuuden ja laadun sekä välimatkan eroavaisuuden' pesyeskeemaksi. Klusterin mahdollinen olemassaolo vahvistaa tutkimushypoteesini ja vastaa laadullisen kandidaatin työni tutkimustuloksia. Erityisesti skeemojen ERO.ERI ja EROTA.ERI käyttäytymisprofiilit muistuttavat toisiaan. Skeema ERO.NUMEROT kytkeytyy yksimuuttujaisen analyysin nojalla ensisijaisesti skeemaan ERO.ERI. Klusteriin B2 ryhmittyy myös transitiivinen EROTTAA.ERI, jonka syntaktinen käytös poikkeavaa monelta osin edellisistä. Sen käyttäytymisprofiili muistuttaa kuitenkin monien ID-merkkien osalta (kuten KVANTTORI ja VASTAUS) intransitiivi- tai kopulauseessa esiintyvien skeemojen käyttäytymisprofiilia. Tulokset ovat kuitenkin alustavia ja erot pieniä. Klusterointi A – jossa samankaltaisuusmittarina on canberran sijaan euklidinen etäisyys – ryhmittää skeeman EROTTAA.ERI skeemojen transitiivisten skeemojen EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA yhteyteen. Skeemoihin ERO.ERI ja EROTA.ERI nähden skeeman EROTTAA.ERI kokonaisfrekvenssi on pieni, mikä sekin heikentää tulosten uskottavuutta. Reliabiliteetin nostamiseksi skeeman EROTTAA.ERI otosta tulisi kasvattaa. Jatkotutkimuksen tehtäväksi jää selvittää, toistuvatko tulokset, mikäli muuttujalukua kasvatetaan tai sovelletaan muita monimuuttujaisia menetelmiä, esimerkiksi regressioanalyysiä.

Toinen tutkimushypoteesieni mukainen ryhmittely on klusterin B3 tai A4 olemassa olo. Tätä klusteria voisi nimittää pesyeskeeman 'havainnointi ja kognitiivinen toiminta' toteutumaksi. Erityisesti lekseemienväliset kytkökset ovat kiinnostavia. Niitä ovat skeemojen EROTTAA.HAVAITA2 ja ERO.YMMÄRTÄÄ sekä ERO.TEHDÄ.ERO ja EROTTAA.LUOKITTELU pariutuminen, joka näkyy sekä A- että B-klusteroinnissa. Skeemat ERO.HAVAITA ja ERO.NÄHDÄ muodostavat lekseeminsäinen parin. B-klusteroinnissa skeema EROTTAA.HAVAITA1 jää sivuun omaksi saarekkeekseen, mutta A-klusteroinnissa se yhdistyy skeemojen EROTTAA.HAVAITA2 ja ERO.YMMÄRTÄÄ joukkoon. Myöhempi yhdistyminen heijastelee, että kaltaisuus on skemaattisempaa. Klusterointi B muistuttaa tältä osin luvussa 5.3.6 tehtyä klusterointia. Tulokset tukevat esimerkiksi luvussa 5.3.5 tekemääni havaintoa, että skeeman EROTTAA.HAVAITA1 "merkityskomponentti" – puhdas aistihavainto – on usein mukana ainakin skeeman EROTTAA.HAVAITA2 spesifioimissa korpusesimerkeissä. Jo klusterianalyyseissä totesin, että skeemojen ERO.TEHDÄ.ERO ja EROTTAA.LUOKITTELU ja sekä skeemojen ERO.YMMÄRTÄÄ ja EROTTAA.HAVAITA2 käyttäytymisprofiilien suhteellinen samankaltaisuus johtuu ensisijaisesti siitä, että *erottaa*-verbin kognitiiviset ja havaintoa ilmaisevat käyttötavat

assosioituvat sellaisten transitiiviverbien kuin *ymmärtää*, *tietää* ja *tehdä* käyttötapoihin – tai konstruktiioihin, joista *ero*-lauseke esiintyy näiden verbien (lause)objektina. Transitiivilauseessa esiintyvät *ero*-nominin skeemat manifestoivat nimenomaan näitä transitiivisia konstruktiota. Niiden kytkös transitiivilauseeseen ja *erottaa*-verbiin onkin ”välillistä”.

Klusteroinnit antavat vahvistusta skeemojen EROTA.IRTI ja EROTA.KIRKKO suhteelliseen samankaltaisuudelle, mitä manifestoi etenkin A-klusteroinnin klusteri A2 [EROTA.IRTI, EROTA.KIRKKO], mutta myös klusteri B1 [[EROTA.KIRKKO ja EROTA.PUOLISO], EROTA.IRTI]. Skeemojen yhteys osoitetaan sekä yksi- että monimuuttujaisen analyysin työkaluilla. Kiinnostavaa on sekin, mitä klusterianalyysi ei löydä, nimittäin skeemojen EROTA.IRTI ja EROTTAA.IRTI Yksimuuttujaisessa analyysissä arveltua yhteyttä, joka ei saa vahvistusta monimuuttujaisesta vaiheesta. Tulos heijastaa sitä tosiasiaa, ettei klusterointi valituilla ID-merkeillä ja asetuksilla halua yhdistää intransitiivisia ja transitiivisia skeemoja.

Yksimuuttujaisen analyysivaiheen tuloksia vahvistaa skeemojen ERO.PUOLISO ja EROTA.PUOLISO yhteys, jota manifestoi etenkin klusteri A3 [ERO.PUOLISO ja EROTA.PUOLISO] ja osaltaan klusteri B1 [[[EROTA.KIRKKO ja EROTA.PUOLISO], EROTA.IRTI], [ERO.IRTI & ERO.PUOLISO]]. Lisäksi yksimuuttujaisessa vaiheessa havaittu yhteys skeemojen EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA välillä saa tukea kummastakin klusteroinnista. Klusteroinnissa A niihin kytkeytyy skeema EROTTAA.ERI ja klusteroinnissa B skeemat EROTTAA.IRTI ja EROTTAA.IRTI2. Yksimuuttujaisen analyysin havaintojen pohjalta luontevammalta, joskaan ei ainoalta oikealta tulkinnalta, tuntuu klusterointiratkaisu A, jossa esiintyy klusterit A5 [ERO.IRTI ja EROTTAA.IRTI], EROTTAA.IRTI2] sekä A6 [EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA], EROTTAA.ERI]. Klusteria A5 voisi luonnehtia ’transitiiviseksi irti-sanomisen tai institutionaalisen suhteen katkaisemisen’ pesyeskeemaksi. Skeemaa A6a [EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA] taas voisi luonnehtia ’abstraktin tai konkreettisen eroamisen skeemaksi’ ja siihen myöhemmin yhdistyvää skeemaa EROTTAA.ERI hylkiöksi, joka liimautuu tähän ryhmään, kun ei muitakaan kavereita ole. Sama ilmiö näkyy luvun 5.3.6 klusterianalyysissä. Klusteroinnissa B sille löytyy semanttisesti sopivampi kaveripiiri.

Keskeisin havaintoni on kuitenkin metodinen. Jo kahdesta eri asetuksilla tehdystä klusteroinnista havaitaan, miten herkästi klusterointialgoritmin tuottama klusteri-struktuuri heittelee (vrt. Glynn 2012: 17). Koska klusteroinnin tulokset ovat selvästi riippuvaisia valituista asetuksista, ei tunnu järkevältä takertua yhteen klusteriratkaisuun ja julistaa sen erinomaisuutta. Sen sijaan tuntuu semanttisesti järkevältä yhdistää A ja B klusteroinnin tuloksia eli poimia eri klusteriratkaisuista semanttisesti mielekkäimmät ryhmittelyt – niin kuin olen-

kin tehnyt. En jatka klusterianalyysin ruotimista tätä enempää, vaan siirryn kuvaaman analyysistä nousevat pesyeskeemat, minkä jälkeen luvussa 8 esitän loppuyhteenvedot.

7 Pesyeskeemojen määrittäminen

Yksimuuttujaisen ja monimuuttujaisen vaiheen tutkimustulokset eivät aseta tiukkoja ehtoja sille, miten skeemojen välisiä, lekseemirajat ylittäviä suhteita tulisi jäsentää. Koska työni on semanttinen, abstrahoin lukujen 1.3 ja 3.3 (ks. myös kuvio 2, s. 20) mallin pohjalta neljä pesyeskeemaa, joissa ryhmäjäsenyys pohjautuu semanttiseen analogiaan. Semanttisella analogialla tarkoitan kahta asiaa. Ensinnäkin kategoriapainotuksia, jotka yhden pesyeskeeman sisällä ovat suhteessa kaltaisempia kuin pesyeskeemojen välillä sekä toiseksi lekseemiskeemojen merkityspotentiaa eli minkälaisia prosesseja tai tiloja ne ilmoittavat.

Pesyeskeemojen määrittelyssä olen käyttänyt apunani paitsi semanttista intuitioita, myös i) luvun 5 yksimuuttujaista analyysivaihetta, ii) luvussa 5.5 abstrahoimiani skeema-kohtaisia kollikatiivisia profiileja sekä iii) etenkin luvun 6 klusteristruktuureja A ja B, joiden tuloksia olen yhdistellyt. Pesyeskeemalla viitataan semanttiseen kokonaishahmoon (vrt. Jääskeläinen 2004: 76, 90), joka on skemaattinen suhteessa johdosten derivaatiostatukseen (ks. luku 2.2.2). Näkökulmani on samantapainen kuin kognitiivisessa kieliopissa (ks. Leino 1993: 80). Siinä johdospesyeen jäsenten ajatellaan edustavan samaa konseptuaalista aluetta, jossa jokainen korostaa ja hahmottaa aluetta omasta (johtimen suuntaamasta) näkökulmasta. Pesyeskeema-näkökulmassa konseptuaalinen alue hahmotetaan ”semeemikohtaisesti” siten, että pesyeen semanttinen kokonaisavaruuksisuus osittuu useaksi limittäiseksi, päällekkäiseksi ja toisiinsa monin tavoin kytkeytyneeksi konseptuaaliseksi alueeksi. Yhdessä ne muodostavat *ero*-pesyeen ”merkitysten verkon” (ks. Lakoff 1987; Langacker 1991: luku 10). Abstrahoimani pesyeskeemat¹⁷³ ovat seuraavat:

¹⁷³ On syytä korostaa, että tämä pesyeskeemaluokitus on vain yksi empiiris-teoreettisesti mahdollinen mallinnus *ero*-pesyeen lekseemirajat ylittävistä, skeemojen välisestä semantiikasta. Kyseessä on idealisointi, ”pelkistys kategorian jäsenten yhteisistä ominaisuuksista” (Onikki 144: 133).

Pesyeskeema 1: 'Ominaisuuden, laadun tai määrän yms. eroavaisuuden skeema'

ERO.ERI, ERO.NUMEROT, EROTA.ERI, EROTTAA.ERI

Pesyeskeema 2: 'Tietämisen, tunnistamisen ja havaitsemisen skeema'

EROTTAA.HAVAITA1, EROTTAA.HAVAITA2, EROTTAA.LUOKITTELU,
ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ, ERO.NÄHDÄ, ERO.HAVAITA

Pesyeskeema 3: 'Irtisanomisen, kirkosta eroamisen, parisuhteen päättymisen ja institutionaalisten suhteiden katkeamisen skeema'

- **alaskeema 3A:** ERO.IRTI, EROTTAA.IRTI, EROTTAA.IRTI2

- **alaskeema 3B:** EROTA.KIRKKO, EROTA.IRTI

- **alaskeema 3C:** EROTA.PUOLISO ERO.PUOLISO

Pesyeskeema 4: 'Abstraktin, sosiaalisen tai konkreettisen erottamisen skeema'

EROTTAA.ABSTRAKTI, EROTTAA.KONKRETIA

Skeemojen luonnetta havainnollistaakseni olen koostanut kullekin "infotaulun". Tauluissa on neljä riviä, joista ensimmäinen sisältää parafrasoin skeeman semanttisesta luonteesta. Toisella rivillä on skeemaan toteumat, lekseemiskeemat. Kolmannella rivillä on A tai B klusteroinnissa pesyeskeemaan kytkeytyvät luokat ja niiden *t*-arvot. Luokista monet ovat suhteellisesti yleisempiä nimenomaisessa pesyeskeemassa kuin muissa. Lihavoituna ovat luokat, jotka tulkintani mukaan ovat ominaisia lähinnä nimenomaiselle pesyeskeemalle. Lihavointi on "höllempää" kuin luvussa 6.2, jossa korostettuna on ainoastaan ne luokat, jotka ovat yhdessä klusterissa selvästi positiivisia ja muissa negatiivisia. Neljännellä rivillä on skeemojen kollikatiiviset profiilit, jotka esittelin luvussa 5.5. A- ja B-klusterointien *t*-arvot eivät ole kaikilta osin vertailukelpoisia, eivät myöskään kollikatiivisten profiilien lihavoinnit, jotka signaloivat herkkyystason 3 havaintoja. Niillä on kuitenkin havainnollistavaa arvoa. Viidennellä rivillä on skeeman esimerkkilauseet.

Taulu 55a kuvaa pesyeskeeman 'Ominaisuuden, laadun tai määrän eroavaisuuden' skeemaa, taulu 55b 'Tietämisen, tunnistamisen ja havaitsemisen' skeemaa, taulu 55c 'Irtisanomisen, kirkosta eroamisen, parisuhteen päättymisen ja institutionaalisten suhteiden katkeamisen' skeema ja taulu 55d pesyeskeeman 'Abstraktin, sosiaalisen tai konkreettisen erottamisen skeema'. Taulukoiden esittämisen jälkeen luon kustakin pesyeskeemasta lyhyen luonnehdinnan. En enää toista muuttujien ja luokkien nimiä, sillä ne lienevät tässä vaiheessa tuttuja (ja jos eivät ole, ks. luku 4).

Taulukko 55a. Pesyeskeema 1: 'Ominaisuuden, laadun tai määrän eroavaisuuden skeema'.

Sanallistus	Ilmaisee Y- ja Z-argumenttien välistä ominaisuuden, laadun, määrän, välimatkan tai muun vastaavan eroavaisuutta.	
Skeeman to- teumat	ERO.ERI, ERO.NUMEROT, EROTA.ERI ja EROTTAA.ERI	
ID-merkkejä klusteroitir- kaisussa B ($t = -1.75$ – $+1.75$)	LUOKKA_KOPULA ($t = +1.14$) PÄÄLUOKKA_AKTIIVI ($t = +0.82$) TEMPUS_PREESENS ($t = +0.99$) MODUS_INDIKATIIVI ($t = +0.83$) VERBI_RAKENNE_ERO.VERBI ($t = +1.18$) MODUS1_NA ($t = +0.81$) MODUS2_NA ($t = +0.91$)	X_SEM_ELOTON ($t = +0.82$) Z_SEM_ELOTON ($t = +0.56$) Y_SEM_ELOTON ($t = +0.60$) AIKA_NA ($t = +0.83$) SY_Y_NA ($t = +0.67$) KVANTTORI_VASTAUS ($t = +1.30$) VASTAUS_LÄSNÄ ($t = +1.29$)
Kollikatiiviset profiilit	ERO.ERI = X.arg {NA_eloton} Y.arg {NP~PRON.RAK_eloton} +KONNEK.+Z.arg {NP_eloton}[+R] EROTTA.ERI = Y.arg {NP~PRON_eloton~konkretia} +REKTIO~KONNEKTORI+Z.arg {NP~PRON_eloton~elollinen} [+R] EROTTAA.ERI = X.arg {NP~PRON.RAK_eloton} Y.arg {NP_joukko} +REKTIO+Z.arg {NP~PRON.RAK_elollinen}	
Esimerkki	ERO: On siinä hiukka ero kuitenkin saako 8v vai elinkautisen. EROTTA: Kannabiksen haitat eivät <u>juuri</u> eroa tupakan tai alkoholin haitoista. EROTTAA: Neron erottaa työntekijästä juuri se asenne, ettei ajattele työtä työnä vaan tekee jotain suurempaa kuin raha.	

Pesyeskeeman 'Ominaisuuden, laadun tai määrän eroavaisuuden skeema' toteumille on ominaista, että sekä Y- että Z-argumentti ovat ilmipantuina siten, että molemmat spesifioi eloton tarkoite. Transitiivilauseessa kyse on kuitenkin elollisten olentojen eroavaisuudesta (ks. s. 132). Lauseet ilmoittavat abstraktin tilan, jossa on usein kyse siitä, että jokin olento jollakin tavalla erilainen kuin jonkin toinen olento. Lauseissa esiintyy ominaisuuden intensiteettiä kvantifioivia adjektiivejä tai adverbeja sekä lausekkeita, jotka ilmoittavat ominaisuuden tai laadun eroavaisuuden proposition, toisin sanoen, mistä ero johtuu. Sen koodaa agentti tai adverbialitädennys. Aikaan tai syyhyn liittyviä adverbialleja tai modaalisia elementtejä ei esiinny. Aikamuotona on preesens, moduksena indikatiivi ja pääluokkana aktiivi. Välttämättömyyttä, kykyä, mahdollisuutta tai mahdottomuutta ilmaisevia modaalisia aineksia ei esiinny. Verbi esiintyy yksin. Pesyeskeemalla, joka vastaa kandidaatin tutkielmani ketjua C 'ominaisuuden poikkeavuus' (ks. luku 1.3), on toteuma kaikissa tutkimissani lekseemeissä.

Tarkastelen seuraavaksi pesyeskeemaa 2, jonka notaatio on 'Tietämisen, tunnistamisen ja havaitsemisen skeema'. Sen "potilastiedot" ovat esillä taulukossa 55b.

Taulukko 55b. Pesyeskeema 2: 'Tietämisen, tunnistamisen ja havaitsemisen skeema'.

Sanallistus	Ilmaisee kahden entiteetin välisen eron havaitsemista, ymmärtämistä, tietämistä tai tunnistamista sekä puhdasta aistihavaintoa.
Toteumat	EROTTAA.HAVAITA1, EROTTAA.HAVAITA2, EROTTAA.LUOKITTELU, ERO.TEHDÄ.ERO, ERO.YMMÄRTÄÄ, ERO.NÄHDÄ, ERO.HAVAITA
ID-merkkejä klusteriratkaisu B:ssä ($t = -1.75 - +1.75$)	<div> LUOKKA_TRANSITIIVI ($t = +0.73$) KIELTO_KIELTEINEN ($t = +1.02$) VERBI_RAKENNE_MUU ($t = +0.62$) RESIPROOKKI_LÄSNÄ ($t = +0.71$) MODUS1_KYKY ($t = +1.12$) SYY_NA ($t = +0.65$) </div> <div> X_MUOTO_PERSON(PERS, NOLLA & INKORPOROITU) ($t = +1.26$) Y_MUOTO_PRON.RAK ($t = +0.95$) Y_SEM_ELOTON ($t = +0.82$) Z_SEM_ELOTON ($t = +0.67$) Z_SEM_KONKRETTIA ($t = +0.55$) </div>
Kollikatiiviset profiilit	HAVAITA1 = X.arg {NP~NOLLA~PERS_eloll.} Y.arg {NP~PRON.RAK_eloton~konkr.} +VAIN_Y+ Z.arg {NP_eloton} HAVAITA2 = X.arg {NP~NOLLA~PERS_eloll.} Y.arg {NP~PRON.RAK_eloton~konkr.} +REKTIO~KONN.+ Z.arg {NP~PRON.RAK_eloton} [+R] LUOKITTELU = X.arg {NOLLA_eloll.} Y.arg {NP_eloton} +KONN.+ Z.arg {NP_eloton} [+R] YMMÄRTÄÄ = X.arg {NP_eloll.} Y.arg {NP_eloton} +KONNEK.+ Z.arg {NP_eloton} [+R] TEHDÄ.ERO = X.arg {PERS_eloll.} Y.arg {NP_eloton} +KONNEK.+ Z.arg {NP_eloton} [+R] HAVAITA = X.arg {NOLLA_eloll.} Y.arg {NP_eloton} +Y_JAOLL.+ Z.arg {NA_eloton} [+R]
Protokorpusesimerkki	EROTTAA.HAVAITA1: Vain joka kymmenes ihminen erottaa karjun hajun. EROTTAA.HAVAITA2: Oppii erottamaan ystävät ja viholliset. EROTTAA.LUOKITTELU: JP, jos kuka, erottaa nämä käsitteet toisistaan. ERO.NÄHDÄ: En minä näe mitään eroa alkoholin ja muiden huumeiden välillä. ERO.TEHDÄ.ERO: Kaiken on nykyään oltava enemmistöä miellyttävää kevytkauraa, ei enää uskalleta tehdä eroa lahjakkuuksien välillä. ERO.YMMÄRTÄÄ: Ranskien ja ranskien kieltäminen ei ole lastensuojelurikos. Se että aiheuttaa pikkuvauvalle kehityshäiriöitä on. Ei luulisi olevan vaikeaa ymmärtää, että näiden välillä on vissi ero. ERO.HAVAITA: Kaikki ruoat onnistuu ilman munaakin, maussa ei huomaa mitään eroa.

Pesyeskeeman 2 toteumat hakeutuvat kielteisiin transitiivilauseisiin, joissa lauseen verbilauseke koostuu usein monesta verbikomponentista. *ero*-lausekkeet voivat jäsentyä infiniittisubjektiksi. Läsnä on usein kykyä (eli kyvyn puutetta tai toiminnan hankaluutta) ilmoittava modaalinen apuverbi tai adjektiivi. Nämä lauseet ilmoittavat tilannetta, jossa ihmistarkoitteinen olento ei kykene toteuttamaan verbin ilmaisemaa propositiota. Agentteina esiintyy elollisia olentoja, jotka NP-subjektin lisäksi koodautuvat pronomineihin tai inkorporoituvat osaksi verbiä tai esiintyvät nollapersoonassa. Sekä Y- että Z-argumentteina esiintyy elottomia tarkoitteita, jotka voivat olla abstrakteja entiteettejä, kuten ajatuksia tai käsitteitä (Lyonsin [1977] 3 asteen entiteettejä), tilanteita (Lyonsin 2 asteen entiteettejä) tai konkreettisia olioita, joilla on spatiaalinen hahmo. Muita pesyeskeemoja useammin tarkoitteet koodautuvat lausetäydennyksiksi. Skeeman toteumat hahmottavat staattisen kognitiivisen tai perceptuaalisen tilan. Skeemat hahmottavat usein kvasiresultatiivisen tilan, jossa objektin sijavaihtelu ei manifestoi resultatiivisuus : irresultatiivisuus -oppositiota, vaan pikemminkin tilan täydellisyyttä : vajaavaisuutta. Skeemassa esiintyy usein resiprookkinen elementti. Skeema vastaa kandidaatin työni merkitysketjua D, 'Aistillinen toiminta' (ks. luku 1.3).

Siirryn nyt tarkastelemaan taulua 55c, jonka notaatio on 'Irtisanomisen, kirkosta eroamisen, parisuhteen päättymisen ja institutionaalisten suhteiden katkeamisen skeema'. Alleviivatut luokat viittaavat kategorioihin, jotka ovat yleisiä useassa alaskeemassa.

Taulukko 55c. Pesyeskeema 3: 'Irtisanomisen, kirkosta eroamisen, parisuhteen päättymisen ja institutionaalisten suhteiden katkeamisen skeema'.

Sanallistutus	Ilmaisee monenlaista eroamista, kuten irtisanomista, kirkosta eroamista, parisuhteen päättymistä ja institutionaalisten suhteiden katkeamista.
Toteumat	alasjeema 3A: ERO.IRT1, EROTTAA.IRT1, EROTTAA.IRT12 – 'transitiivinen eroaminen' alasjeema 3B: EROTA.KIRKKO, EROTA.IRT1 – 'intransitiivinen eroaminen' alasjeema 3C: EROTA.PUOLISO ERO.PUOLISO – 'parisuhteen päättyminen'
ID-merkit klusteri- ratkaisun A perus- teella ($t = -1.79 - +2.25$)	<div>3A 'transitiivi'</div> <div>X_MUO.NA.TRAN. ($t = +1.69$)</div> <div><u>Z SEM INSTT. ($t = +1.64$)</u></div> <div><u>Y SEM INSTT. ($t = +1.46$)</u></div> <div><u>MODUS_VÄLT. ($t = +1.41$)</u></div> <div><u>SY Y LÄSNÄ ($t = +1.41$)</u></div> <div><u>PÄÄLUOK. _PASS. ($t = +1.37$)</u></div> <div><u>TEMPUS PERF. ($t = +1.12$)</u></div> <div><u>Y_MUOTO_NP ($t = +1.01$)</u></div> <div><u>AIKA LÄSNÄ ($t = +1.00$)</u></div> <div><u>Y SEM YKSILÖ ($t = +0.95$)</u></div> <div><u>TEMPUS MUU ($t = +0.77$)</u></div> <div><u>X SEM INSTT. ($t = +0.72$)</u></div> <div><u>MODUS MUU ($t = +0.60$)</u></div>
	<div>3B 'intransitiivi'</div> <div><u>LUOKKA INTR. ($t = +1.91$)</u></div> <div><u>Y MUOTO_PERS ($t = +1.88$)</u></div> <div><u>Y SEM YKSILÖ ($t = +1.74$)</u></div> <div><u>Z SEM INSTT. ($t = +1.68$)</u></div> <div><u>MODUS2 MAHD. ($t = +1.68$)</u></div> <div><u>X MUOTO_NA ($t = +1.46$)</u></div> <div><u>AIKA LÄSNÄ ($t = +1.37$)</u></div> <div><u>MODUS MUU ($t = +1.09$)</u></div> <div><u>TEMPUS MUU ($t = +1.01$)</u></div> <div><u>VERB.VER.KETJU ($t = +0.88$)</u></div> <div><u>SY Y LÄSNÄ ($t = +0.86$)</u></div> <div><u>RESIPROOKKI_NA ($t = +0.79$)</u></div> <div><u>KVANTTORI_NA ($t = +0.64$)</u></div> <div><u>Y MUOTO_NA ($t = +0.50$)</u></div>
	<div>3C 'puoliso'</div> <div><u>Y SEM JOUKKO ($t = +2.35$)</u></div> <div><u>Y MUOTO_NA ($t = +1.83$)</u></div> <div><u>Z MUOTO_NA ($t = +1.80$)</u></div> <div><u>Z SEM ELOLL. ($t = +1.24$)</u></div> <div><u>X MUOTO_NA ($t = +1.10$)</u></div> <div><u>TEMPUS PERF. ($t = +1.04$)</u></div> <div><u>LUOKKA INTR. ($t = +1.02$)</u></div> <div><u>SY Y LÄSNÄ ($t = +1.01$)</u></div> <div><u>RESIPROOKKI_NA ($t = +0.8$)</u></div> <div><u>MODUS1_NA ($t = +0.69$)</u></div> <div><u>AIKA LÄSNÄ ($t = +0.68$)</u></div> <div><u>MODUS2 MAHD. ($t = +0.65$)</u></div> <div><u>TEMPUS MUU ($t = +0.64$)</u></div> <div><u>Y SEM YKSILÖ ($t = +0.40$)</u></div>
Kollika- tiiviset profiilit	<p>EROTA.PUOLISO = Y.arg {NP~NOLLA_joukko} +Y_JAO LLINEN~NA+ Z.arg {NA_elollinen}</p> <p>ERO.PUOLISO = X.arg {NA} Y.arg {NA_elollinen} +NA+ Z.arg {NA}</p> <p>KIRKKO = Y.arg {NP~NOLLA~INKORP. _yksilö~joukko} +REKTIO~VAIN_Z+ Z.arg {NP_instit.}</p> <p>EROTA.IRT1 = Y.arg {NP~PERS_yksilö~instit.} +REKTIO~VAIN_Y+ Z.arg {NP~NA_instituutio}</p> <p>ERO.IRT1 = X.arg {NA} Y.arg {NP_elollinen} +REKTIO~VAIN_Y+ Z.arg {NA_elollinen}</p> <p>EROTTAA.IRT1 = X.arg {NA_eloll.} Y.arg {NP_yksilö~joukko} +VAIN_Y+ Z.arg {NA_instit.}</p> <p>IRT12 = X.arg {NA} Y.arg {NP_instit.} +REKTIO~KONN.+ Z.arg {NP_instit.} [+R]</p>
Protokor- pus esi- merkki	<p>ERO.PUOLISO = Oleskeluluvan saatuaan tulee ero, ja puoliset jäävät veronmaksajien vastuulle.</p> <p>EROTA.PUOLISO = Lesbota eroaa siinä kuin heterot, mutta homot ei.</p> <p>EROTA.KIRKKO = Kun tuln täysi-ikäiseksi erosin heti kirkosta ja lähdin merille .</p> <p>EROTA.IRT1 = Erosin armeijasta, ja suosittelen kaikkia harkitsemaan sitä käsi sydämellä.</p> <p>ERO.IRT1 = Lähes kaksi promillea aikuisväestöstä vaatii ministerin eroa.</p> <p>EROTTAA.IRT1 = Soini ei erottanut halla-ahoa.</p> <p>EROTTAA.IRT12 = Jokohan vanhoillislestadiolaisuus erotetaan kansankirkosta?</p>

Pesyeskeema 3 koostuu kolmesta alasjeemasta, jotka ovat 3A 'transitiivisen eroamisen skeema', 3B 'intransitiivisen eroamisen skeema' ja 3C 'parisuhteesta eroamisen skeema'. Pesyeen alasjeemat ilmoittavat fyysisen, sosiaalisen tai institutionaalisen suhteen päättymistä. Kaikille alasjeemoille yhteistä on esiintyä menneessä aikamuodossa eli imperfektissä tai pluskvamperfektissä (skeemat 3A ja 3C lisäksi preesensissä). Yhteistä on myös aikaa tai syytä ilmoittavien adverbiaalien läsnä olo. Lisäksi yhteistä on, että Y-argumentin spesifioi yksilötarkoitteinen elollinen olento. Skeemat ilmoittavat perfektisen prosessin (tai sen mahdollisuuden), jonka toteutuessa tosimaailmassa tapahtuu muutos.

Alaskeeman 3B 'intransitiivisen eroamisen skeema' lähteeksi hakeutuu instituutiotar-koitteisia olentoja. Agentiksi hakeutuu etenkin yksilöitä, mutta myös joukkoja ja instituutioita, jotka NP-subjektin rinnalla esiintyvät nollapersoonassa, verbiin inkorporoituna tai persoonapronominissa. Lauseessa on usein modaalista mahdollisuutta tai mahdottomuutta ilmaiseva sana. Alaskeemat 3A ja 3B vastaavat kandidaatin työni merkitysketjua B, 'Työsuhteen tai vastaavan päättyminen', ja sen sisällä intransitiivisuus : transitiivisuus -jatkumoa. Alaskeeman 3C 'parisuhteen päättymisen skeema' agenteiksi hakeutuu joukkotarkoitteisia olentoja ja Z-argumenteiksi elollisia tarkoitteita. Yleensä Z-argumentti jää kuitenkin ilmipanamatta ja myös Y-argumentti on usein poissa rakenteesta. Se vastaa kandidaatintyöni (ks. luku 1.3) merkitysketjua A, 'Parisuhteen katkeaminen'. Alaskeemoista 3A 'transitiivisen eroamisen skeema' hakeutuu passiiviin. Sekä Y- että Z-argumentteina esiintyy instituutiotarkoitteisia olioita ja Y-argumentteina lisäksi yksilötarkoitteisia olentoja. Tempus saattaa olla perfektikin ja modus muu kuin indikatiivi eli käytännössä konditionaali. Lauseissa esiintyy välttämättömyyteen liittyviä modaalisia sanoja. Useat piirteet kytkevät alaskeemalla 3C kohti pesyeskeemaa 4, johon skeemat EROTTAA.IRT1 ja EROTTAA.IRT2 sijoittuvat klusteriratkaisussa B. Skeemoille 3C ja 4 yhteistä on muun muassa taipumus passiiviin ja sekä, että agenttina on instituutiotarkoitteinen olento. Tarkastelen lopuksi pesyeskeema 4 taulukon 55d avulla. Sen notaatio on 'Abstraktin, sosiaalisen tai konkreettisen erottamisen skeema'.

Taulukko 55d. Pesyeskeema 4: 'Abstraktin, sosiaalisen tai konkreettisen erottamisen skeema'¹⁷⁴.

Sanallistus	Ilmoittaa abstraktia, sosiaalista tai konkreettista eroamista
Toteumat	EROTTAA.ABSTRAKTI, EROTTAA.KONKRETIA (EROTA.ABSTRAKTI, EROTA.KONKRETIA)
ID-merkkejä klusteriratkaisun A perusteella ($t = -1.79 - +2.25$)	Y_SEM_KONKRETIA ($t = +1.00$) Z_MUOTO_NP ($t = +0.81$) PÄÄLUOKKA_PASSIIVI ($t = +0.78$) PAT_MUOTO_NP ($t = +0.76$) KIELTO_NEUTRAALI ($t = +0.75$) LUOKKA_TRANSITIIVI ($t = +0.75$) X_MUOTO_NA.TRANS ($t = +0.74$) VERBI_RAKENNE_MUU ($t = +0.60$) Z_SEM_KONKRETIA ($t = +0.50$) X_SEM_INSTITUUTIO ($t = +0.50$)
Kollikatiiviset profiilit	ABSTRAKTI = X.arg {NA_eloton} Y.arg {NP_eloton} +REKTIO + Z.arg {NP_eloton} [+R] KONKRETIA = X.arg {NA_eloll.} Y.arg {NP_konkr.} +REKTIO + Z.arg {NP_konkr.} [+R]
Protokorpus-esimerkki	ABSTRAKTI: Suomi tunnustaa Ahvenanmaan itsenäisyyden ja erottaa sen itsestään. KONKRETIA: Tarkoituksena on erottaa virtsa ulosteista.

¹⁷⁴ Skeeman ID-merkkien t -arvot olen seulonut klusterista A6[[EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA], EROTTAA.ERI] siten, että olen poistanut t -arvot, jotka arvioni mukaan johtuvat skeeman EROTTAA.ERI vaikutuksesta.

Pesyeskeema 4 ilmoittaa abstraktin suhteen päättymistä tai konkreettista fyysistä manipulaatiota. Skeemojen EROTTAA.ABSTRAKTI ja EROTTAA.KONKRETIA isoin ero onkin merkitystausta, jota vasten ne hahmottavat predikaation. Jälkimmäisessä se on fyysinen, edellisessä abstrakti. Pesyeskeeman korpuksimerkit hakeutuvat passiivilauseisiin ja välttelevät kieltolauseita. Y- ja Z-argumentteina esiintyy elottomia tarkoitteita, konkreettisia tai abstrakteja entiteettejä. Agentteina esiintyy sekä elottomia että elollisia, fyysiseen manipulaatioon kykeneviä olentoja. Agentteina esiintyy myös instituutioita. Y- ja Z-argumentit koodautuvat luokkaan NP, mikä tarkoittaa, että niiden rooleissa ei esiinny pronomineja tai lausetäydennyksiä. Se myös tarkoittaa, että molemmat argumentit ovat ilmipantuina. Pesyeskeema 4 vastaa yksimuuttujaisessa vaiheessa luvussa (5.3.5) erittelemääni passiivihakuista tyyppiä.

Lopetan pesyeskeemojen esittelyn tähän. Tähän loppuu myös tutkimustulosteni esittäminen. Olen kahdessa analyysivaiheessa – yksimuuttujaisessa ja monimuuttujaisessa analyysissa – antanut ainakin välttävät vastaukset johdannossa esittämääni kolmeen tutkimuskysymykseen, jotka käännettynä vastauksiksi ovat: i) olen selvittänyt, että määrällinen analyysi tukee hypoteesia siitä, että *ero*-lekseemit ovat monimerkityksisiä, ii) olen kuvannut *ero*-skeemojen polysemian sekä lekseemitasolla johdospesyeen tasolla sekä iii) paikantanut muuttujat, jotka kytkeytyvät skeemojen merkitysvaihteluun lekseemien sisällä ja välillä.

On loppusanojen aika.

8 Lopuksi

Johdantoluvussa esitin pro gradu -työni päätavoitteeksi selvittää kvantitatiivisten työkalujen avulla, miten *ero*-, *erota*- ja *erottaa*-lekseemien monimerkityksisyys rajauksien puitteissa jäsentyy paitsi kunkin lekseemin sisäisessä merkitysavaruudessa myös *ero*-lekseemien välillä johdospesyeen tasolla. Tätä tarkoitusta varten poimin Helsingin Sanomien uutiskomentointiaineistosta (HS.fi-korpus) 1666 korpuskatkelman kokoelman, joka käsittää 509 *erottaa*-verbin, 563 *erota*-verbin ja 594 *ero*-nominin sisältävää esimerkkikatkelmaa. Koodasin jokaiselle korpuksimerkille tiedon siitä, mitä muuttujan alakategoriaa se edustaa suhteessa 19 muuttujaan. Muuttujat on esitelty ja määritelty luvussa 4. Lisäksi koodasin jokaiselle esimerkille tiedon siitä, mihin luvussa 3 määrittelemistäni skeematyypeistä se lukeutuu. *erottaa*-verbille määritin 9, *erota*-nominille 7 ja *ero*-lekseemille 8 pääskeemaa.

Luvussa 5 suoritan kolme erillistutkimusta, yhden kullekin lekseemille. Näissä analyysissä tarkastelen yksimuuttujaisten menetelmien skeemojen ja X-, Y- ja Z-argumenteiksi

nimittämieni rooliargumenttien välistä assosiaatiota. X-argumentti vastaa työssäni transitiivilauseen agenttia. Y-argumentti vastaa transitiivilauseen patienttia, intransitiivilauseen agenttia tai kopulalauseessa esiintyvän *ero*-nominin ensimmäistä täydennystä. Z-argumentti vastaa transitiivi- ja intransitiivilauseessa rektioadverbiaalia, jota nimitän lähteeksi. Toisinaan Z-argumentti vastaa rinnasteisen Y-argumentin *ja*-merkin tai vastaavan jälkeen tulevaa jaksoa. Kopulalauseessa Z-argumentti koodaa *ero*-nominin toista täydennystä. Argumentteja vastaavat muuttujat ovat X_MUOTO ja X_SEM (X-argumentti), Y_MUOTO ja Y_SEM (Y-argumentti) sekä Z_MUOTO ja Z_SEM (Z-argumentti). Yksimuuttujaisessa vaiheessa tutkin lisäksi muuttujia RAKENNE ja RESIPROOKKI, joista jälkimmäinen mittaa *ero*-lausekkeen täydennyksenä esiintyvää resiprookkista elementtiä, jonka yleisimmät spesifikaatiot ovat *toisistaan*- ja *välillä*-sanat. Muuttuja RAKENNE tarkastelee *ero*-lausekkeiden argumenttirakennetta. Etenkin *ero*-nominin osatutkimuksessa hyödynsin analyysissä tukimuuttujia.

Yksimuuttujainen analyysivaihe osoittaa, että skeemojen ja muuttujien alaluokkien välillä on systemaattista kytköksisyyttä, joka palautuu muun muassa siihen, esiintyykö ilmira-kenteessa ilmipantuja X, Y, ja Z-argumentteja sekä toisaalta siihen, mitä semanttista kategoriaa ilmipannut rooliargumentit edustavat. Analyysin yhteenvedoissa kullekin skeemalle määritellään yksilöllinen kollikatiivinen profiili, joka hahmottelee, miten yksimuuttujaisessa vaiheessa mukana olleet muuttujaluokat tyypillisesti esiintyvät kussakin skeemassa. Profiilit vihjaavat, että skeemojen välillä on myös lekseemirajat ylittävää analogisuutta.

Yksimuuttujainen analyysivaihe osoittaa lisäksi, että ensisijaisesti *ero*-verbien analyysiin luomani muuttujisto soveltuu paremmin *ero*-verbien kuin *ero*-nominin semantiikan erittelyyn. Esimerkiksi verbin argumenttirakennetta mittaamaan luotu muuttuja RAKENNE ei vangitse *ero*-nominin merkityksenvaihtelua toivotulla tavalla. Positiivinen löytö on esimerkiksi se, että myös *ero*-nominilla on elatiivirektiota skeemassa ERO.IRT1 (*Rusasen mukaan ero ensisijaisesta hoivaajasta voi aiheuttaa lapsessa ahdistusta ja hylkäämisen tunteita*). Luokan REKTIO semanttiseksi sukulaissieluksi *ero*-nominissa paljastuu REKTIO2, jossa *ero*-sijaisen (elatiivin spesifioiman) relaation sijaan entiteettien välinen suhde koodautuu tulosijaisella ”rektiolla”. Näiden tyyppien sukulaisuutta demonstroi minimipari *pölynimurin ero hiustenkuivaajaan on tosi iso :pölynimuri eroaa hiustenkuivaajasta tosi paljon*. Keskeisemmässä roolissa *ero*-nominin ”argumenttirakenteen” vaihtelussa on *ero*-nominin rakennetyypit, joista yleisimmät esittelin luvussa 4.1.2. Rakennetyypit heijastavat muun muassa erilaisia ”kopulalausekonstruktioita” (ks. s. 47 alav. 51), jollaisia havainnollistaa pari *Kanan ja lehmän ero on se, että toisesta tulee maitoa ja toisesta munia : Navettaeläimissä ei ole mi-*

tään eroa. Verbin ja nominin välinen sijamuotojen vertailu on ongelmallista, sillä verbitäydennykset esiintyvät muutamassa kielipillisessä sijassa. Tässä heijastuu, miten hankalaa on luoda muuttujisto, joka on yhtä aikaa tehokas sekä verbin että nominin tarpeisiin.

Johdospesyeen määrällisen koeasetelman suunnittelu ja toteutus sisältääkin ongelmia, koska siinä kontrastoidaan kielen kategorioita, joiden syntaktinen käyttäytyminen poikkeaa perustavanlaatuisella tavalla. Koska tiedossani ei ole ollut ennakkotapauksia koskien johdospesyeen semanttisista analyysiä määrällisin menetelmin, olen joutunut nyhjäisemään tyhjistä koodaussysteemin, josta halusin luoda sellaisen, että aineistoa voidaan analysoida käyttäytymisprofilein. Ratkaisuksi johdosten luontaiseen ”epäsymmetriaan” soveltamani X-, Y- ja Z-argumentteihin pohjaava rooliargumenttien jaottelu on parhaimmillaankin välttävä. Vastaus viimeiseen (neljanteen) tutkimuskysymykseeni on, että johdospesyeen polysemiaa on mahdollista tarkastella valitsemillani määrällisillä analyysivälineillä mielekkäitä tuloksia saaden. Vastausta ei kuitenkaan pidä ymmärtää lopulliseksi ratkaisuksi. Klusterianalyysissä (luvussa 6) olen tukimuuttujilla (esim. LUOKKA ja NA.TRANS) pyrkinyt paikkaamaan luokitussysteemiini vinoumia, esimerkiksi sitä, että Y-argumentti hahmottaa transitiiivilauseessa patientin, intransitiivilauseessa agentin ja kopulalauseessa usein genetiivimääritteen. Tällaisten tukimuuttujien käyttö voi kuitenkin johtaa siihen, että luvussa 6 soveltamani klusterianalyysi alkaa ryhmitellä sanoja lekseemikohtaisesti, jolloin tuloksena on kuin toisinto sanakirjasta, joka paljastaa lekseemien sisäisen struktuurin, muttei lekseemien välistä struktuuria, jonka paljastus on työni päätavoitteita. Käyttäytymisprofileihin syötettävissä muuttujissa on pyritty löytämään kultainen keskite: toisaalta ei saa kertoa klusterianalyysille, että tutkittavat sanat ovat syntaktisesti täysin erilaisia, muttei myöskään saa vääristää lingvististä totuutta siten, että jätetään ilmiön kannalta olennaisia muuttujia sivuun.

Käyttäytymisprofileita on sovellettu enimmäkseen englannin kielen synonyymisten tai polyseemisten lekseemien tarkasteluun. Esimerkiksi Gries (2006) ja Glynn (2012) tutkivat *run*-sanon monimerkityksisyyttä. Arppe (2008), joka ensisijaisesti käyttelee regressioanalyysiä, tutkii ”mieltä”-sanojen synonymiaa. Synonymiaa tai polysemiaa tutkivien koeasetelmien järjestäminen onkin helppoa, sillä tyypillisesti niitä sentään sitoo sanaluokkanalogia. Johdospesyeen tutkimuksessa yhdistyy tavallaan sekä synonymian tutkimus että monimerkityksisyyden tutkimus, mikä tekee siitä jännittävää mutta ”menetelmällisesti” painajaismaista. Tässä työssä soveltamani koeasetelma, jolla tarkoitan muuttujistoa, mutta myös analyysimenetelmiä, ei ole lopullinen, vaan vaatii monenlaista kehittämistä.

Luvussa 6 analysoin skeemakohtaisia käyttäytymisprofileja monimuuttujaisella me-

netelmällä, jonka nimi on agglomeratiivinen hierarkkinen klusterianalyysi. Klusterianalyysissä tarkasteltavana ovat kaikki aineistolle koodaamani muuttujat, muuttujaa RAKENNE lukuun ottamatta. Klusterianalyysin tulokset ovat sikäli kannustavia jatkotutkimusta silmällä pitäen, että analyysi paljastaa hypoteesieni mukaisia kytköksiä *ero*-pesyeen skeemojen välillä. Tulokset ovat kuitenkin alustavia muun muassa useiden skeemojen riittämättömän frekvenssin vuoksi.

Valitsin klusterianalyysin analyysimenetelmäksi sillä perusteella, että useat korpuslingvistiikkaan suuntautuneet ja kognitiivisesti orientoituneet tutkijat pitävät sitä työkaluna, joka sopii polyseemisten sanan eri merkitysten (*senses*) tai synonyymisten sanojen välisten suhteiden erittelyyn (ks. esim. Glynn 2012: 18). Klusterianalyysin etuja on, ettei siinä aseteta tiukkoja ehtoja muuttujien alakategorioiden jakaumille. Gries (2010b: 22) huomauttaa, ettei klusterianalyysi ole lyönyt itseään samalla tavalla läpi korpuslingvistiikassa kuin tietokone-lingvistiikassa, jossa sen käyttö on yleistä. Hän arvelee karsastuksen johtuvan toisaalta siitä, että klusterianalyysin tuottamia puudiagrammeja on pidetty jollain tavoin mystisinä ja toisaalta siitä, että sekä puudiagrammeja tuottavien koeasetelmien järjestäminen että niiden tulkitseminen edellyttää teknistä harjaantumista. Ehkä näin onkin. Klusterianalyysin ja yksimuuttujaisten *post hoc* -testauksien soveltamista varten jouduin itekin syventymään *R*-ohjelmointikieleen, jota käytetään yleisesti korpusorientoituneessa kielentutkimuksessa (ks. esim. Arppe 2008, Baayen 2008, Gries 2003, Glynn 2012). Koska *R* on suunniteltu nimenomaan tilastotieteilijöiden (ei ohjelmoijien) tarpeisiin, se on ohjelmointikielenä yksinkertainen ja helposti omaksuttava, eikä se edellytä aikaisempaa ohjelmointitaitausta. Lisäksi se on ilmainen. Kaupallisiin ohjelmiin (kuten ihmistieteissä suosittuun SPSS:ään, ks. Nummenmaa 2004) verrattuna *R*:n etu lingvistin näkövinkkelistä katsottuna on, että *muutkin lingvisti käyttävät sitä*. Helposti saatavilla on vertaistutkimuksia, oppaita ja lingvistien tarpeisiin tehtyjä skriptejä, esimerkiksi klusterianalyysii tai *post hoc* -testausta varten (Arppe 2015).

Tarkkojen *post hoc* -testauksien tarpeellisuutta pohdin luvussa 5.5. Tulen lopputulokseen, että niillä on lisäarvoa muuttujien ja skeemojen välisten yhteyksien paljastamisessa. Usein tosin jo prosenttijakauma paljastaa ilmiön luonteen. (Agglomeratiivinen) hierarkkinen klusterianalyysi sisältää monia ongelmakohtia. Se on toimintaperiaateiltaan kuin ”hölmöläinen”, joka uskoo kaiken mitä sille kerrotaan. Tällä tarkoitan, että klusterointialgoritmi suorittaa klusteroinnin orjallisesti tutkijan tekemien valintojen pohjalta. Lopputulokseen vaikuttaa etenkin se i) mitä samankaltaisuusmittaria käytetään, ii) mitä yhdistämissääntöä käytetään, iii) miten alakategorioita käyttäytymisprofiileihin syötetään ja toisaalta miten aineisto on koodattu sekä iv) kuinka monta klusteria klusteristruktuurissa valitaan olevan. Manning

& Schütze (1999: 499) nostavat yhdeksi hierräkkisen klusterianalyysin menetelmälliseksi ongelmaaksi lisäksi sen, että algoritmin täytyy valita yksi klusteri johon jokainen klusteri (skeema) ryhmitetään. Jokainen skeema pakotetaan kuulumaan johonkin klusteriin, vaikka klusterin skeemoilla ei olisi mitään luonnollista kielitieteellistä yhteyttä. Klusterointialgoritmi ei voi nostaa käsiään ilmaan ja todeta, että jokin klusteri (skeema) on niin erilainen, ettei sitä kannata ryhmitellä mihinkään klusteriin. Klusterianalyysi on myös herkkä poikkeaville havainnoille. Luotettavien klusterointitulosten saamiseksi olisikin suoritettava huolellinen yksi- ja kaksimuuttujainen esitutkimus.

Työni puutteita on se, etten suorita kaksimuuttujaisista vaihetta. Myös työn yksimuuttujainen vaihe (luku 5) on sikäli riittämätön, että analysoin vain 8 muuttujaa. Lisäksi on syytä korostaa, että rajasin työstäni pois joukon ilmiön kannalta todennäköisesti relevantteja kategorioita. Poissaolollaan loistavat muuttujat, jotka mittaavat verbin lukua (*erosin, erosit, erosimme*), lauseen syntaktista statusta (päälause, sivulause, kysymyslause jne.), objektin ja nominin ”rajattuutta” : ”täydellisyyttä” (verbissä totaaliobjekti : partitiiviobjekti -oppositio, nominissa nominatiivi : partitiivi -oppositio). Lisäksi sivuun jäi kategorioita, jotka mittaavat *ero*-lekseemeissä yleisiä adverbialityyppejä, jotka ilmoittavat välinettä (*kodeiini erotetaan kahvinkeittimellä*), erottamisperustetta (*Itse erotan hajun perusteella tupakan sähkötupakasta satavarmasti;*), keinoa tai reittiä (*Voisiko työpaikastakin erota ilmoittamalla jollekin välityssivustolle*), kehystä (*Omassa perheessäni ainakin ero oli lapsille helpotus ja palautti rauhan*; *Oliko kyselyssä erotettu romaanikerjäläiset ja suomalaiset romaanit?*) tai ”fokusta” (*En näe tämän asian suhteen eroa saatananpalvojien, juutalaisten, muslimien tai kristittyjen välillä*). On hankala arvioida, miten tällaisten muuttujien läsnäolo muuttaisi esimerkiksi klusterointituloksia. Hypoteesini on, että ne selkeyttäisivät niitä, sillä etenkin adverbialiluokilla ja skeemoilla on havaintojeni mukaan jokseenkin idiosynkraattisia kytköksiä.

Jatkotutkimuksen tehtäväksi jääkin i) ratkoa klusterianalyysiin liittyvät metodologiset ongelmat, ii) ulottaa yksimuuttujainen vaihe kaikkiin muuttujiin, iii) selvittää luokkien väliset kaksimuuttujaiset kytkökset ja iv) oikaista työn monimuuttujaisien tulosten mahdolliset vääristymät. Vaikkei tämän tutkimuksen monimuuttujainen vaihe olisikaan täysin moitteeton mallitapaus, se toiminee silti (varoittavana) esimerkkinä siitä, kuinka vastaavanlaisia (ja myös hyvin erilaisia) koe-asetelmia voidaan järjestää ja tutkia.

Työssäni on muitakin mahdollisia ongelmakohtia. Objektivisuuden näkökulmasta ongelma saattaa olla, että olen itse i) luonut ja määritellyt muuttujat, niiden alakategoriat ja *ero*-lekseemien skeemat sekä ii) koodannut ja iii) analysoinut aineiston. Jatkotutkimusta silmällä pitäen jonkinlainen vallan kolmijako voisi olla paikallaan. Olen luokkien määrittelyssä

ja aineistoni koodauksessa kuitenkin tietoisesti pyrkinyt mahdollisimman suureen objektiivisuuteen ja läpinäkyvyyteen sekä siihen, etten ratkaisullani johdattelisi klusterianalyysia hypoteesieni mukaiseen lopputulokseen. Klusterianalyysia on nimittäin helppo kuljettaa kuin pässiä narussa, jos paimen näin haluaa tehdä.

Oma menetelmällinen ongelmavyöhyke on aineistoesimerkkien annotointi skeemoiksi, joka tässä työssä on intuitiivista. Luvussa 3.1 hahmotan Kilgarriffin (2003) nojaten näkökulmia siitä, että polyseemisen sanan ”eri merkityksiä” voitaisiin esiryhmitellä esimerkiksi klusterianalyysin avulla. Mutta senkin onnen edessä on esteitä. On hankala kuvitella, että mikään suomeen sovellettavissa oleva algoritmi olisi yhtä luotettava merkitysseula kuin harjaantuneen kielentutkijan suorittama merkityksentunnistus (vrt. Itkonen 2010: 157), tosin siinäkin on ongelmansa. Divjak ja Gries (2009:6) kritisoivat tietokone-lingvistikä korpus-tutkimusta (jollaista esimerkiksi Arppen 2008 tutkimus on) siitä, että ne luottavat liikaa semi-automaattisiin prosessointityökaluihin, jotka tuottavat melua (*noise*) ja vääristävät tutkimustuloksia. Parhaimpiin lopputuloksiin päästäneekin, kun ihminen liittoutuu koneen kanssa tai kun empirismi liittoutuu intuition kanssa. Tähän ajatukseen pohjautuu luvussa 7 suorittamani pesyeskeemojen määrittely, jossa yhdistelen semanttiseen tietooni skeemojen merkityseroista kahden eri klusterointiratkaisun eli A- ja B-klusterointien tuloksia sekä lisäksi yksimuuttujaisen analyysin tuloksia. Tällaista menetelmää voidaan kritisoida puhtaasti tilastotieteellisestä näkökulmasta. Toisaalta kielentutkimusta, eritoten semantiikkaa, ei voida redusoida tilastotieteeksi (vrt. Karlsson yms. 1980: 94). Puhtaasti semanttisesta näkökulmasta yksi työn kehityspiste on se, että noudatan liiankin tunnontarkasti X^2 -testauksen miniehdotuksia (vrt. Arppe 2008: 101–103, joka ottaa vapauksia miniehdotuksista). Seuraus on luokkien yhdistelystä johtuva informaatiokato, joka karkeistaa tutkittavaa ilmiötä.

Edellä mainituista ongelmista huolimatta työni tarjoamat tutkimustulokset – joita raportoin lukujen 5 ja 6 yhteenvedoissa, luvussa 7 sekä analyysissä läpi työn – osoittavat, että *ero*-pesyeen skeemojen tai alamerkitysten välillä on eroavaisuutta, joka voidaan palauttaa käyttäytymisprofileihin (monimuuttujainen analyysi) ja kollokatiivisiin profileihin (yksimuuttujainen analyysi). Työ siis paljastaa, että *ero*-kantaisen sanojen merkityksenvaihtelua voidaan palauttaa *ero*-sanojen seuralaisten semanttisiin ominaisuuksiin tai jopa hyvin yksinkertaiseen vaihteluun siinä, onko jokin seuralainen ilmipantuna vai ei. Lisäksi työ paljastaa, että pesyeestä on abstrahoitavissa karkeasti ajateltuna 4 pesyeskeema, jotka voi luonnehtia parafraseilla: 1) ’ominaisuuden, laadun, määrän tai vastaavan eroavaisuuden skeema’, 2) ’tietämisen, tunnistamisen tai havaitsemisen skeema’, 3) ’irtisanomisen, kirkosta eroamisen,

parisuhteen päättymisen ja institutionaalisten suhteiden katkeamisen skeema' sekä 4) 'abstraktin, sosiaalisen tai konkreettisen eroamisen skeema'. Tulokset ovat kaikessa riemastuttavuudessaan alustavia: analyysi rajoittuu HS.fi-korpukseen, valitsemiini lekseemeihin, tehtyihin rajauksiin ja tutkimuksessa mukana olleisiin muuttujiin.

Lopuksi on syytä pohtia, tuoko määrällinen analyysi merkittävää lisäarvoa intuitiopohjaiseen korpustutkimukseen, jota edustaa esimerkiksi *ero*-pesyettä tarkasteleva kandidaatintyöni. Kokonaisuudessaan tulokset ovat samansuuntaisia kuin kandidaatintyössä, mikä tarkoittaa, että *ero*-sanojen merkitysvaihtelun pääpiirteet voidaan vangita intuition avulla. Khiin neliö -testien minirajoitusten vuoksi yksimuuttujainen analyysi jopa hämärtää *ero*-pesyeen kokonaiskuva: analyysistä joudutaan laittamaan sivuun paitsi "Zipfianilaista" distribuutiota manifestoivia (vrt. s. 86 alav. 100), korpuksessa harvinaisia skeemoja eli *ero*-sanojen käyttötapoja myös niukkafrekvenssisii luokkia, joilla on skeemojen kanssa yksilöllisiä konfiguraatioita. Intuitiopohjaisen analyysin vahvuus on nimenomaan kiinnostavien poikkeavuuksien kuvauksessa ja toisaalta ilmiön kokonaiskuvan piirtämisessä.

Yksimuuttujainen analyysi, joka tarkastelee pelkästään kategorioiden frekvenssieroja, paljastaa kuitenkin merkitysvaihtelun tendenssejä, joiden havaitseminen kieli-intuition avulla – korpusta käytellen tai käyttelemättä – on hankalaa. Yksimuuttujainen analyysi antaa esimerkiksi vahvaa näyttöä siitä, että parisuhteesta eroamisen merkitys on käsitteistys, jossa Y-argumentti (se kuka eroaa) että etenkin Z-argumentti (se mistä tai kenestä erotaan) jäävät yleisesti ilmipanematta. Toisaalta kirkosta eroamisen merkitys edellyttää ilmipantua Z-argumenttia, joka spesifioi, että kyse on nimenomaan uskonnollisen kytköksen päättymisestä. Rektiotäydennyksen ilmipanemisen vaade vihjaa, että kyseessä ei ole "itsenäinen merkitys". Frekvenssipohjainen analyysi antaa toisin sanoen luotettavaa tietoa siitä, mitkä tiettyyn skeemaan *potentiaalisesti* kytkeytyvistä (intuition avulla pääteltävissä olevista tai yksittäistapauksien nojalla havaituista) kielenominaisuuksista *aktuaalisesti* esiintyvät skeeman yhteydessä. Lisäksi se kertoo, onko piirre yleinen vai harvinainen paitsi skeeman sisällä myös skeemojen välillä. Koska *erottaa*-verbillä esimerkiksi nollapersoonaa on huomattavasti harvinaisempi irtisanomista (EROTTA.IRTI) kuin havaitsemista tai kategorisoimista (EROTTA.HAVAITA1–2, EROTTAA.LUOKITTELU) ilmoittavien skeemojen yhteydessä, voidaan päätellä, nollapersoonaa-agentti on edelliselle pikemminkin poikkeus, kun taas jälkimmäisille yleinen agentin ilmaisustrategia.

Useita muuttujia analysoidaan vain monimuuttujaisessa vaiheessa, joskin jo niiden yksimuuttujainen analyysi paljastaisi, että skeemoilla on systemaattisia eroja esimerkiksi siinä, mitä aikamuotoja ne työllistävät: menneet aikamuodot ovat yleisiä etenkin parisuhteen tai

työsuhteesta päättymisen merkityksissä. Tällaiset kytkökset jäävät helposti huomaamatta intuitiopohjaisessa analyysissä huomion kiinnittyessä ”kiinnostavimpiin” piirteisiin. Kandidaatintyössäni huomaamatta ovat jääneet sellaisetkin seikat, että resiprookkista täydennystä tai konstruktiota, jossa erotettavat entiteetit kytketään toisiinsa *ja*-konnektorilla, esiintyy vain tietyillä skeemoilla. Esimerkkejä olisi lukuisia.

Työn klusterointiratkaisut eivät sinänsä tuota paljoa uutta tutkimustietoa sen päälle, mitä tiesin *ero*-pesyeen johdostenvälisistä suhteista kandidaatintyön pohjalta. Klusteroinnin merkittävin tieteellinen anti on ehkä sen osoittamisessa, että suomen kielen sanojen monimerkityksisyyttä voidaan suhteellisen hyvin tuloksin tutkia käyttäytymisprofileihin pohjautuvalla lähestymistavalla. On hankala arvioida, mikä hyöty tästä on puhtaasti semantiikan analyysin kannalta, jos kerran samansuuntaisia tuloksia saadaan ilman työlästä määrällistä koejärjestelyä. Mitä enemmän tarkasteltavia kielenpiirteitä on, sitä hankalammaksi intuitiopohjainen analyysi kuitenkin käy. Sitä suuremmaksi käy riski, että tutkija luo virheellisiä tulkintoja päätellessään, mitkä ominaisuudet ovat ilmiön kannalta relevantteja ja mitkä eivät ole. Klusterianalyysissä sovelletut *t*-arvot ilmoittavat juuri tätä: miten kielenpiirteet painottuvat klusterien välillä.

Työni kirvoittaa joitakin jatkotutkimushankkeita. Yksimuuttujaisen analyysien yhteenvedoissa sivuan automaattisen merkityksentunnistamisen algoritmiiikkaa. Voitaisiin tutkia, onko kollikatiivisten profiilien antamalla tiedolla sovellusarvoa esimerkiksi kieliteknologian saralla polysemian identifioimisessa tai semiautomaattisen kielenkääntämisen sovellutuksissa. Semantiikassa tarvitaan tapauskohtaisia tutkimuksia, jotka paljastavat kielen semanttisen ja rakenteellisen puolen ominaisuuksia. Tämä tapaustudkimus paljastaa, että *ero*-pesyeessä on abstrahoitavissa merkityskeskittymiä, skeemoja, jotka manifestoivat yksilöllisiä, lekseemiskeema- ja pesyeskeemakohtaisissa ko(n)tekstuaalisia ja kollikatiivisia konfiguraatioita. Laadullisessa kandidaatintutkielmassani tekemät havainnot *ero*-pesyeen semanttisen rakenteen ontologista saavat empiiristä todistusaineistoa.

Työni kirvoittaa pohtimaan, minkälaisiin suomen kielen ilmiöihin käyttäytymisprofileihin perustuvaa lähestymistapaa voitaisiin soveltaa – mielestäni monenlaisiin ilmiöihin. Koska käyttäytymisprofileihin voidaan inkorporoida tietoa kielen semanttisista, morfologisista, syntaktisista, pragmaattisista – ja ekstralingvistisistä ominaisuuksista, se ei aseta tutkimuskohteilla rajoja. Käyttäytymisprofileihin soveltuvaa lähestymistapaa voikin soveltaa lähes mihin tahansa kielen ilmiöön, olettaen, että hyväksytään hypoteesi, että kielen ilmiöiden erot palautuvat eroihin ilmiöiden tai ominaisuuksien esiintymistaajuuksissa. Tällaisia kohteita ovat (lähi)synonyymit ja monimerkityksiset sanat (tai jopa polyseemiset sanueet, kuten

tässä työssä), erilaiset grammaattiset rakenteet, murteentutkimus, keskusteluntutkimus, tekstintutkimus tai miksei jopa kirjallisuudenkin tutkimus.

Siinä kaikki.

Lähteet

- AGRESTI, ALAN 2002: *Categorical Data Analysis*. 2. painos. Hoboken: John Wiley & Sons.
- ARPPE, ANTTI 2008: *Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy*. Helsinki: Helsingin yliopisto. Saatavissa: http://www.ling.helsinki.fi/~aarppe/Publications/Arppe_Dissertation_Final_Full.
- ARPPE, ANTTI 2015 [2013]: *Polytomous-ohjelmapaketti R-kielille*. Edellyttää vähintään versiota R 2.10.1. Saatavissa: <https://cran.r-project.org/web/packages/polytomous/index.html>.
- ARPPE, ANTTI & JÄRVIKIVI, JUHANI 2007: Every method counts - Combining corpus-based and experimental evidence in the study of synonymy. – *Corpus Linguistics and Linguistic Theory* 3/2 s. 131-159.
- ATKINS, BERYL T. 1987: Semantics ID Tags: Corpus evidence for dictionary senses. The Uses of Large Text Databases. *Third Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary* s. 17–36. Waterloo, Canada.
- BAAYEN, R. HARALD 2001: *Word Frequency Distribution*. Dordrecht: Kluwer Academic Publishers.
- BAAYEN, R. HARALD 2008: *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. New York: Cambridge University Press.
- BACKHAUS, KLAUS – ERICHSON, BERND – PLINKE, WULFF – WEIBER, ROLF 2003: *Multivariate Analysemethoden: eine anwendungsorientierte Einführung* [Monimuuttujaiset analyysimenetelmät: käytännön opas]. 10. painos. Berlin & Heidelberg & New York: Springer.
- BIBER, DOUGLAS – CONDRAD, SUSAN – REPPEN, RANDI 1998: *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- BIBER, DOUGLAS & JONES, JAMES K. 2005: Merging corpus linguistic and discourse analytic research goals: Discourse units in biology research articles. – *Corpus linguistics and linguistic theory* 1/2 s. 151–182. London: De Gruyter.
- BOLINGER, DWIGHT L. 1968: Entailment and the meaning of structures. – *Glossa* 2 s. 119–127.
- BULGAKOV, MIHAIL 2008 [1969]: *Saatana saapuu Moskovaan*. Ulla-Liisa Heino (suom.). Helsinki: WSOY.
- CORBETT, MICHAEL & LE ROY, MICHAEL 2002: *Research Methods in Political Science: an Introduction Using Microcase*. Belmont: Wadsworth Publishing.
- COCHRAN, WILLIAM G. 1954: Some Methods for Strengthening the Common X^2 Test. – *Biometrics Vol 10/4* s. 417–451.
- COHEN, JACOB 1992: A Power Primes – *Psychological Bulletin* 1/112 s. 155–159.
- CORNISH, ROSIE 2007: *Cluster analysis*. [Verkkodokumentti]. Mathematics Learning Support Centre. Saatavissa: <http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf>. [Viitattu 2.10.2015].
- CRUSE, D. ALAN 1986: *Semantics*. Cambridge: Cambridge University Press.
- DENISON, NORMAN 1957: *The Partitive in Finnish*. Helsinki: Suomalaisen tiedeakatemian toimituksia B 108.

- DESHORS, SANDRA C. 2010: *A multifactorial study of the uses of may and can in French-English interlanguage*. Brighton: Department of Linguistics. University of Sussex. Saatavissa vain sähköisesti: <http://sro.sussex.ac.uk/6987/>.
- DIVJAK, DAGMAR 2010: *Structuring the Lexicon: a Clustered Model for Near-Synonymy*. Cognitive Linguistics Research 43. Berlin: De Gruyter.
- DIVJAK, DAGMAR & FIELLER, NICK 2014: Cluster analysis: Finding structure in linguistic data. Teoksessa Glynn, Dylan & Robinson, Justyna (toim.) *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy* s. 405–441. Amsterdam & Philadelphia: John Benjamins.
- DIVJAK, DAGMAR & GRIES, STEFAN TH. 2006: Ways of trying in Russian: clustering behavioral profiles. – *Corpus Linguistics and Linguistic Theory* 2/1 s. 23–60.
- 2008: Clusters in the mind. Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3:2/2008 s. 188–213.
- 2009: Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. Teoksessa Evans, Vyvyan & Pourcel, Stephanie S. (toim.) *New directions in cognitive linguistics* s. 57–75. Amsterdam & Philadelphia: John Benjamins.
- 2009b: Corpus-based cognitive semantics: A contrastive study of phrasal verbs in English and Russian. Teoksessa Dziwirek, Katarzyna & Lewandowska-Tomaszczyk, Barbara (toim.) *Studies in cognitive corpus linguistics* s. 273–296. Frankfurt am Main: Peter Lang.
- EVANS, VYVYAN & GREEN, MELANIE 2006: *Cognitive Linguistics*. Edinburgh: Edinburgh University Press Ltd.
- EVERT, STEFAN & KRENN, BRIGITTE 2001: Methods for the qualitative evaluation of lexical association measures. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics* s. 188–195. Toulouse, France.
- FIRTH, JOHN RUPERT 1968 [1957]: A synopsis of linguistic theory 1930–55. Teoksessa Palmer, Frank Robert (toim.) *Selected Papers of J. R. Firth 1952–59*. London: Longmans.
- GEERAERTS, DIRK 2006: Introduction. A rough guide to cognitive linguistics. Teoksessa Geeraerts, Dirk (toim.) *Cognitive linguistics. Basic readings* s. 1–28. Berlin: Mouton de Gruyter.
- GILDEA, DANIEL & JURAFSKY, DANIEL 2001: Automatic labeling of semantic roles. *Computational Linguistics* 28/3 s. 245–288. Saatavissa: <http://www.mitpressjournals.org>.
- GOLDBERG, ADELE E. 1995: *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- GOODMAN, LEO A. & KRUSKAL, WILLIAM H. 1954: Measures of Association for Cross-Classifications. – *Journal of the American Statistical Association* 49/268 s. 732–764.
- GLYNN, DYLAN 2012: The many uses of run. Corpus-based methods and socio-cognitive semantics. Teoksessa Glynn, Dylan & Robinson, Justyna A. (toim.) *Corpus Methods in Cognitive Semantics*. Amsterdam: John Benjamins. Saatavissa: http://www.dsglynn.univ-paris8.fr/articles/Glynn_2012_Polysemy_run.pdf

- 2014: Correspondence Analysis. Exploring data and identifying patterns. Teoksessa Glynn, Dylan & Robinson, Justyna (toim.) *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy* s. 443–486. Amsterdam & Philadelphia: John Benjamins.
- GRIES, STEFAN TH. 2003: *Multifactorial analysis in corpus linguistics: a study of particle placement*. London: Continuum.
- 2006: Corpus-based methods and cognitive semantics: the many meanings of to run. Teoksessa Gries, Stefan Th. & Stefanowitsch, Anatol (toim.) *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis* s. 57–99. Berlin & New York: Mouton de Gruyter. Griesin artikkelit saatavissa: <http://www.linguistics.ucsb.edu/faculty/stgries/research/overview-research.html>. [Vitätattu 10.10.2015].
- 2010. *Behavioral Profiles 1.01*. Ohjelma R-kielelle. Edellyttää vähintään versiota R 2.7.1. Saatavissa tekijältä pyydettäessä.
- 2010b: Useful statistics for corpus linguistics. Teoksessa Sánchez, Aquilino & Moisés, Almela (toim.) *A Mosaic of corpus linguistics: selected approaches* s. 269–291. Frankfurt am Main: Peter Lang.
- 2012: Behavioral profiles. A fine-grained and quantitative approach in corpus-based lexical semantics. Teoksessa Jarema, Gorja – Libben, Gary – Westbury, Chris (toim.) *Methodological and analytic frontiers in lexical research* s. 57–80. Amsterdam & Philadelphia: John Benjamins.
- 2013: *Statistics for linguistics with R*. 2. painos. Berlin & New York: De Gruyter Mouton.
- 2015a (tulossa): Quantitative methods in linguistics. Teoksessa Wright, James D. (toim.) *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Elsevier.
- 2015b: Some current quantitative problems in corpus linguistics and a sketch of some solutions. – *Language and Linguistics 16/1* s. 93–117. Saatavissa: http://www.ling.sinica.edu.tw/files/publication/j2015_1_5_3464.pdf.
- 2015c (tulossa): Quantitative designs and statistical techniques. Teoksessa Biber, Douglas & Reppen, Randi (toim.) *The Cambridge Handbook of Corpus Linguistics*. Cambridge: Cambridge University Press.
- GRIES, STEFAN TH. & OTANI, NAUKI 2010: Behavioral profiles: a corpus-based perspective on synonymy and antonymy. – *ICAME Journal 34* s. 121–150.
- GRIES, STEFAN TH. & STOLL, SABINE 2009: How to measure development in corpora? An Association strength approach. – *Child Language 36/2009* s. 1075–1090. Cambridge University Press: United Kingdom.
- GRONDELAERS, STEF – GEERAERTS, DIRK – SPEELMAN, DIRK 2007: A case for a cognitive corpus linguistics. Teoksessa Gonzalez-Marquez, Monica – Mittelberg, Irene – Coulson, Seana – Spivey, Michael J. (toim.) *Methods in Cognitive Linguistics. Human Cognitive Processing 18*. Amsterdam: John Benjamins.
- HANKS, PATRICK 1996: Contextual dependency and lexical sets. – *International Journal of Corpus Linguistics 1:1/2006*, s. 75–98.
- HARRIS, ZELLIG S. 1970: *Papers in structural and transformational linguistics*. Dordrecht: Reidel.
- HAKULINEN, AULI & KARLSSON, FRED 1979: *Nykysuomen lauseoppi*. Helsinki: SKS.

- HEIKKILÄ, TARJA 2008: *Tilastollinen tutkimus*. 7. uudistettu painos. Helsinki: Edita.
- HOLOPAINEN, MARTTI & PULKKINEN, PEKKA 2003: *Tilastolliset menetelmät*. Helsinki: WSOY.
- HOEY, MICHAEL 2005: *Lexical Priming: A new Theory of Words and Language*. London: Psychology Press.
- HOWELL, DAVID 1999: *Fundamental Statistics for the Behavioral Sciences*. 4. painos. Pacific Grove: Brooks & Cole Publishing Company.
- HS.FI-UUTISKOMMENTOINTIAINEISTO [Helsingin Sanomien Uutiskomentointiaineisto]
- BORIN, LARS – FORSBERG, MARKUS – ROXENDAL, JOHN 2012: *Korp – the corpus in frastructure of Språkbanken*. Saatavissa luvanvaraisesti: korp.csc.fi. URN:NBN:FI:LB-2014052717. [Viitattu 6.6.2015].
- HUUMO, TUOMAS 2001: Kvasiresultatiivilauseet: esimerkki dynaamisesta subjektiivisesta käsitteistyksestä – *Virittäjä* 105 s. 85–101.
- 2006: Kvantiteetti ja aika I. Nominaalisen aspektin näkökulma suomen objektin ja subjektin sijamerkintään. – *Virittäjä* 100 s. 504–538. Saatavissa: http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/2006_504.pdf.
- 2006b: Näkökulmia suomen kielen aistihavaintoverbeihin. – *Emakeele Seltsi aastaraamat* 52/2006 s. 69–86. Saatavissa: www.emakeeleselts.ee/esa/ESA_52_pdf/Huumo.pdf.
- HUUMO, TUOMAS & SIVONEN, JARI 2010: Leonard Talmy – kognitiivisen kielitieteen elävä klassikko. Haddington, Pentti & Sivonen, Jari (toim.) *Kielentutkimuksen modernit klassikot – kognitiivinen ja funktionaalinen kielitiede* s. 19–40. Helsinki: Gaudeamus.
- HÄKKINEN, KAISA 2007: *Nykysuomen etymologinen sanakirja*. Helsinki: WSOY.
- ITKONEN, ESA 1997: The social Ontology of Linguistic Meaning. Teoksessa Haukioja, Timo – Helasvuo, Merja-Liisa – Miestamo, Matti (toim.) *Suomen kielitieteellisen yhdistyksen vuosikirja 1997* s. 50–80. Turku: Suomen kielitieteellinen yhdistys.
- 2010: *Maailman kielten erilaisuus ja samuus. Osa III*. Turun yliopisto. Yleisen kielitieteen julkaisuja.
- ITKONEN, ESA & PAJUNEN, ANNELI 2010: *Empiirisen kielitieteen metodologia*. Helsinki: SKS.
- ITKONEN, TERHO 1976: Erään sijamuodon ongelmia. Teoksessa *Opuscula instituti linguae Fennicae Universitas Helsingiensis* 53 s. 173–217. Helsinki: Helsingin yliopisto.
- JANDA, LAURA & SOLOVYEV, VALERY 2009: What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS. – *Cognitive Linguistics* 20/2. 367–393. London: De Gruyter.
- JANTUNEN, HARRI JARMO 2009: Ei pelkästään mielikuvituksen puutteen vuoksi – Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virittäjä* 113 s. 101–113.
- JOHNSON, MARK 1987: *The Body in the mind: The bodily basis of Meaning, Imagination, And Reason*. Chicago: University of Chicago Press.
- JÄÄSKELÄINEN, PETRI 2004: *Instrumentatiivisuus ja nykysuomen verbinjohto. Semanttinen tutkimus*. Jyväskylä Studies in Humanities 22. Jyväskylä: Jyväskylän yliopisto.

- KANGASNIEMI, HEIKKI 1992: *Modal expressions in Finnish*. Helsinki: Suomalaisen kirjallisuuden seura.
- KARLSSON, FRED – HAKULINEN, AULI – VILKUNA, MARIA 1980: *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Helsingin yliopiston yleisen kielitieteen laitoksen julkaisuja 6. Helsinki: Helsingin yliopisto.
- KAUFMAN, LEONARD & ROUSSEAU, PETER J. 1990: *Finding Groups in Data*. New York: John Wiley.
- KETCHEN, DAVID J. – SHOOK, CHRISTOPHER, L. 1996: *The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique*. – *Strategic Management Journal* 17/6 s. 441–458.
- K1–1666 = Tolan 2015 koostama tutkimuskorpus. Saatavissa pyydettäessä.
- KILGARRIFF, ADAM 2003: “I don’t believe in word senses”. Teoksessa Nerlich, Brigitte – Todd, Zazie – Herman, Vimala – Clarke D. David (toim.) *Polysemy. Flexible Patterns of Meaning in Mind and Language* s. 361–392. Berlin: Mouton de Gruyter. Saatavissa: <http://www.kilgarriff.co.uk/Publications/1997-K-CHum-believe.pdf>.
- 2005: Language is never, ever, ever random. – *Corpus Linguistics and Linguistic Theory* 1–2/2005 s. 263–275. Saatavissa: <https://kilgarriff.co.uk/Publications/2005-K-linear.pdf>
- KISHNER, JEFFERY M. & GIBBS, RAYMOND W. 1997: How just gets its meanings: Polysemy and context in psychological semantics. – *Language and Speech* 39/1 s. 19–36.
- KTOO = KANKAANPÄÄ, SALLI (toim.) – HEIKKILÄ, ELINA – KORHONEN, RIITTA – MAAMIES, SARI – PIEHL, AINO 2009: *Kielitoimiston oikeinkirjoitusopas*. 6. painos. Kotimaisten kielten tutkimuskeskuksen julkaisuja 147. Helsinki: Kotimaisten kielten tutkimuskeskus.
- KOIVISTO, VESA 1991: *Suomen verbikantaisten UtU-verbijohdosten semantiikkaa*. Suomi 161. Helsinki: SKS.
- KYTÖMÄKI, LEENA 1977: *Suomen verbijohdosten generointia*. Lisensiaatintyö. Turku: Turun yliopiston suomalaisen ja yleisen kielitieteen laitos.
- KS = *Kielitoimiston sanakirja. Osat 1–3*. 2006. Helsinki: Kotimaisten kielten tutkimuskeskus. (Myös elektronisena.)
- LAITINEN, LEA 1995: Nollapersoonaa. – *Virittäjä* 99 s. 337–358.
- LAAKSO, JOHANNA 1990: *Translatiivinen verbinjohdin NE itämerensuomalaisissa kielissä*. SUST 204. Helsinki: Suomalais-Ugrilainen Seura.
- LAKOFF, GEORGE 1987: *Women, fire and dangerous things*. Chicago: The University of Chicago. Saatavissa: <http://emilkirkegaard.dk/en/wp-content/uploads/George-Lakoff-Women-Fire-and-Dangerous-Things.pdf>.
- LAKOFF, GEORGE & JOHNSON, MARK 1980: *Metaphors we live by*. Chicago: University of Chicago Press.
- LANGACKER, RONALD W. 1987: *Foundations of cognitive grammar I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- 1991: *Foundations of Cognitive Grammar II: Descriptive Application*. Stanford: Stanford University Press.
- 1999: *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

- 2008: *Cognitive Grammar – A Basic Introduction*. Oxford: Oxford University Press.
- LEHTINEN, TAPANI 1979: *Itämerensuomen verbien historiallista johto-oppia: suomen avajaa, karkajaa -tyyppiset verbit ja niiden vastineet lähisukukielissä*. SUST 169. Helsinki: Suomalais-Ugrilainen Seura.
- LEINO, JAAKKO 2005: *Luento 4: Valenssista* [verkkodokumentti]. Julkaistu 5.10.2005. Saatavissa: www.helsinki.fi/~jaaleino/kielioppi2/. [Viitattu: 1.10.2015].
- LEINO, PENTTI 1991: *Lauseet ja tilanteet: Suomen objektin ongelmia*. Suomi 160. Helsinki: SKS.
- 1993: *Kieli 7. Polysemia – kielen moniselitteisyys*. Helsinki: Helsingin yliopiston suomen kielen laitos.
- LNA = Lönnrotiana-kokoelma 103 II: *Suomen kielen johto-oppia*. Suomalaisen Kirjallisuuden Seuran Kirjallisuusarkisto.
- LUOJOLA, TEEMU 2006: *Kielitieteellisen aineiston kvantitatiiviset analyysimenetelmät*. Helsinki: Helsingin yliopiston yleisen kielitieteen laitos. Saatavissa: www.ling.helsinki.fi/~fkarlss/methods/kvant_men.
- LYONS, JOHN 1977: *Semantics 1–2*. Cambridge: Cambridge University Press.
- LÖFBERG, LAURA – ARCHER, DAWN – PIAO, SCOTT – RAYSON, PAUL – MCENERY, TONY – KRISTA, VARANTOLA – JUNTUNEN, JUKKA-PEKKA 2003: Porting an English semantic tagger to the Finnish language. Teoksessa: Dawn, Archer– Rayson, Paul – Wilson, Andres – McEnery, Tony (toim.) *Proceedings of the Corpus Linguistics 2003 conference. UCREL technical paper number 16* s. 457–464. Lancaster: UCREL. Saatavissa: <http://ucrel.lancs.ac.uk/publications/CL2003/papers/lofberg.pdf>.
- MANNING, CHRISTOPHER D. & SCHÜTZE, HINRICH 1999: *Foundations of Statistical Natural Language Processing*. Cambridge: The MIT Press. Saatavissa: http://ics.upjs.sk/~pero/web/documents/pillar/Manning_Schuetze_StatisticalNLP.pdf
- MILLER, GEORGE A. 1990: Nouns in WordNet: a lexical inheritance system. – *International Journal of Lexicography* 3/4 s. 245–264. Saatavissa: <ftp://ftp.cogs.ci.princeton.edu/pub/wordnet/5papers.ps>
- MORAN, MATTHEW D. 2003: Arguments for rejecting sequential Bonferroni in ecological studies. – *OIKOS* 100/2, s. 403–405. Saatavissa: http://people.stfx.ca/rscrosat/biology384/moran_03.pdf.
- NS = *Nykysuomen sanakirja I–IV*. 1951–1961. Helsinki: WSOY.
- NUMMENMAA, LAURI 2004: *Käyttätymistieteiden tilastolliset menetelmät*. Helsinki: Tammi.
- OJUTKANGAS, KRISTA 1997: Suomen ja viron U- ja ne-verbijohdosten käytön vertailua. – *Virittäjä* 101 s. 358–380.
- OJUTKANGAS, KRISTA 2008: Kielentutkijan mielikuvituksen puute eli mihin todelliseen kielenkäyttöön perustuvaa aineistoa tarvitaan. – *Virittäjä* 112 s. 276–284. Saatavissa: http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/2008_276.pdf.
- ONIKKI, TIINA 1994: Skeema merkitysrakenteen ja taustatiedon kuvauksessa. Teoksessa Leino, Pentti & Onikki, Tiina 1994 (toim.) *Kieli 8. Näkökulmia polysemiaan*. Helsinki: Helsingin yliopiston suomen kielen laitoksen julkaisuja.

- ONIKKI-RANTAJÄÄSKÖ, TIINA 2001: *Sarjoja. Nykysuomen paikallissijaiset olotilanilmaukset kielen analogisuuden ilmentäjinä*. Suomalaisen Kirjallisuuden Seuran Toimituksia 817. Helsinki: SKS.
- PAJUNEN, ANNELI 1988: *Verbien leksikaalinen kuvaus*. Helsingin yliopiston yleisen kielitieteen laitoksen julkaisuja 18. Helsinki: Helsingin yliopisto.
- 2001: *Argumenttirakenne. Asiantilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Helsinki: SKS.
- PENTTILÄ, AARNI 1963 [1957]: *Suomen kielioppi*. 2. painos. Helsinki: WSOY.
- PITKÄNEN, ANTTI J. & KOHONEN, VILJO 1984: *Johdatus kvantitatiiviseen kielentutkimukseen ja alan atk-sovelluksiin*. Helsinki: Gaudeamus.
- PS = *Suomen kielen perussanakirja 1–3*. 1990–94. Helsinki: Valtion painatuskeskus.
- RAUKKO, JARNO 1994: Polysemia: Merkitysten verkosto – Merkityksen verkko. Teoksessa Leino, Pentti & Onikki, Tiina (toim.) *Kieli 8. Näkökulmia polysemiaan* s. 36–69. Helsinki: Helsingin yliopiston suomen kielen laitos.
- R-OHJELMA = R-tietokoneohjelma. Versio R 3.1.2. Saatavissa: <http://www.r-project.org/>.
- ROSCH, ELEANOR & LLOYD, BARBARA B. 1978: *Cognition and categorization*. Hillsdale: Erlbaum.
- ROSNOW, ROBERT & ROSENTHAL, RALPH A. 1989: Statistical Procedures and the Justification of Knowledge in Psychological Science. – *American Psychologist* 44 / (10) s. 1276–1284. Saatavissa: http://ist-socrates.berkeley.edu/~maccoun/PP279_Rosnow.pdf
- RÄISÄNEN, ALPO 1978: Kantasanan ja johdoksen suhteesta. – *Virittäjä* 82 s. 321–344.
- RÄISÄNEN, ALPO 1988: *Suomen kielen u-johdimiset verbit*. Suomi 141. Helsinki: SKS.
- SALMELA, JUSSI & KOHONEN, VILJO 1977: CHITAB – A ”Poor Man’s” Shortcut to Computer Processing of Linguistic Data. Teoksessa Gellerstam, Martin (toim.) *Nordiska datalingvistikdagar 1977. Rapporter från Språkdata*. Sverige: Göteborgs universitet.
- SALMINEN, JUTTA 2012: Epäilen, että tämä verbi kyseenalaistaa täydennyksensä proposition. *Epäillä*-verbin merkityksen kehityksestä ja monitulkintaisuudesta. Pro gradu -tutkielma. Helsingin yliopiston suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos. Saatavissa: <https://helda.helsinki.fi/handle/10138/29618.5>
- SETÄLÄ, EMIL NESTOR 1966 [1880]: *Suomen kielen lauseoppi*. 11. painos. Tarkastanut Matti sadeniemi. Helsinki: Otava.
- SINCLAIR, JOHN 1998: The Lexical Item. Teoksessa Weigand, Edda (toim.) *Contrastive Lexical Semantics* s. 1–24. Amsterdam/Philadelphia: John Benjamins.
- SINHA, CHRIS & KUTEVA, TANIA 1995: Distributed spatial semantics. – *Nordic Journal of Linguistics* 19 s. 167–199.
- SSA = *Suomen sanojen alkuperä. Etymologinen sanakirja. I–III*. Helsinki: Suomalaisen Kirjallisuuden Seura & Kotimaisten kielten tutkimuskeskus 1992–2000.
- SUZUKI, RYOTA & SHIMODAIRA, HIDETOSHI 2014 [2006]: Pvcust: an R package for Hierarchical clustering with p-values. – *Bioinformatics* 22/12 s.1540–1542. Saatavissa: <http://www.sigmath.es.osaka-u.ac.jp/shimo-lab/prog/pvcust/>.

- SULLIVAN, GAIL M. & FEINN, RICHARD 2014: Using effect size – or why the P value is not enough. – *Journal of Graduate Medical Education* 3/4 s. 279–282.
- TALMY, LEONARD 1985: Lexicalization Patterns: Semantic Structure in Lexical Forms. Teoksessa Shopen, Timothy (toim.) *Language Typology and Syntactic Description, Vol III: Grammatical Categories and the Lexicon* s. 57–149. Cambridge: Cambridge University Press.
- 2005: Foreword. Teoksessa Gonzalez-Marquez, Monica – Mittelberg, Irene – Coulson, Seana – Spivey, Michel J. (toim.) *Methods in Cognitive Linguistics* s. XI–XXI. Amsterdam & Philadelphia: John Benjamins.
- THEIL, HENRI 1970: On the Estimation of Relationships Involving Qualitative Variables. – *American Journal of Sociology* 1/76 s. 103–154.
- TILASTOKESKUS 2015: *Käsitteet ja määritelmät* [verkkodokumentti]. Saatavissa: <http://www.stat.fi/meta/kas/tilastoyksikko.html>. [Viitattu 1.10.2015].
- TOLA, JOONATAN 2010: Kumpi johtaa, kumpi seuraa johtajaa. Kandidaatintutkielma. Helsingin yliopiston suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos.
- TOGNINI-BONELLI, ELENA 2001: *Corpus linguistics at work*. Philadelphia: John Benjamins.
- 2002: Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. Teoksessa Altenberg, Bengt & Granger, Sylviane (toim.) *Lexis in Contrast: Corpus-based approaches* s. 73–95. Amsterdam: Benjamins.
- TUGGY, DAVID 2006: Schematic network. Ambiguity, polysemy, and vagueness. Teoksessa Geeraerts, Dirk (toim.) *Cognitive linguistics: Basic readings* s. 167–184. New York: Mouton de Gruyter.
- TUGGY, DAVID 2007: Schematicity. Teoksessa Geeraerts, Dirk & Cuyckens, Hubert (toim.) *The Oxford Handbook of Cognitive Linguistics* s. 80–116. New York: Oxford University Press.
- TUMMERS, JOSE – HEYLEN, KRIS – GEERAERTS, DIRK 2005: Usage-based approaches in cognitive linguistics: A technical state of the art. – *Corpus linguistics and linguistic theory* 1/2 s. 225–261. Berlin: Walter de Gruyter.
- VISK 2004 = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA-RIITTA – ALHO, IRJA: *Iso suomen kielioppi*. SKST 950. Verkkoaversio. Saatavissa: <http://scripta.kotus.fi/visk>. URN:ISBN:978-952-5446-35-7. [Viitattu 1.10.2015].
- VILKUNA, MARIA 1980: *Referenssi ja spesies suomenkielisten tekstien tulkinnassa*. Lisensiaatintyö. Helsinki: Helsingin yliopiston suomen kielen laitos.
- 1996: *Suomen lauseopin perusteet*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 90. Helsinki: Edita.
- WARD, JOE H. JR. 1963: Hierarchical groupings to optimize an objective function. – *Journal of the American Statistical Association* 58/301 s. 236–244.
- WIKIPEDIA = Artikkelit Vs [verkkodokumentti]. Saatavissa: <https://fi.wikipedia.org/wiki/Vs>. [Viitattu 1.10.2015].
- WITTGENSTEIN, LUDWIG 1981: *Filosofisia tutkimuksia*. Heikki Nyman (suom.). Juva: Taskutieto 155.

YBUNGALOBILL = Nimimerkki verkkokeskustelussa *Cluster analysis in R: determine the optimal number of clusters* [verkkodokumentti]. 2012. Saatavissa: <http://stackoverflow.com/questions/15376075/cluster-analysis-in-r-determine-the-optimal-number-of-clusters>. [Viittauspäivä 10.10.2015].

Liite 1: Herkkyystaulukoita

erota-verbin taulukoita

Taulukko 25x. Muuttujan Y_MUOTO X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 24.995. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI		IRTI		KIRKKO		PUOLISO	
NP	+5.4	4.3	+0.1	+0.6	-9.7	-5.2	-0.2	-0.8
NOLLA	-16.0	-5.5	0.0	+0.1	+13.7	+4.5	+3.9	+2.3
INKORPOR.	-10.8	-4.4	0.0	+0.2	+17.7	+5.0	0.0	0.0
PERS	-9.6	-4.0	+5.3	+2.8	+1.7	+1.5	0.0	+0.1
NA	-7.0	-3.4	+2.3	+1.8	+0.2	+0.5	+2.6	+1.8
PRON	+45.5	+9.0	-14.7	-4.8	-6.6	-3.0	-5.6	-2.7

Taulukko 26x. Solukohtainen X^2 -jakauma (tummalla taustalla) ja standardoidut Pearsonin jäämät (vaalealla taustalla). Herkkyystaso 1 = solu > 2, ja herkkyystaso 3 = solu > 24.996.

	ERI		IRTI		KIRKKO		PUOLISO	
ELOTON	+15.1	+76.96	-8.0	-30.46	-8.6	-35.08	-1.8	-1.9
INSTITUUTIO	-15.9	-60.22	+9.5	+30.0	+10.1	+33.84	-3.6	-5.5
ELOLLINEN	-2.3	+2.70	-3.2	-7.02	-3.2	-7.17	+10.0	+90.8

Taulukko 27x. Muuttujan Z_MUOTO X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < x. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI		IRTI		KIRKKO		PUOLISO	
NA	+7.3	+5.4	-2.8	+3.2	+7.4	+4.9	-29.7	-9.3
NP	-18.8	-6.7	+4.3	+3.0	-8.2	-4.0	+60.7	+10.2
PRON_RAK	+2.8	+2.2	+0.2	+0.5	-3.4	-2.1	-1.2	-1.2

Taulukko 28x. Muuttujan RAKENNE X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 24.995. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI		IRTI		KIRKKO		PUOLISO	
REKTIO	+15.5	+7.1	-2.2	-2.5	+0.5	+1.2	-28.1	-8.1
KONNEKTORI	+20.4	+5.8	-4.6	-2.6	-4.6	-2.5	-2.9	-1.9
VAIN_Z	-13.5	-4.9	+0.8	+1.2	+22.6	+5.7	-1.4	-1.4
Y_JAOLLINEN	+0.3	+0.7	-9.2	-3.8	-12.7	-4.2	+66.4	+9.3
NOLLA	-10.6	-4.2	-0.1	-0.4	0.0	0.0	+35.4	+6.6
VAIN_Y	-20.7	-6.3	+31.4	+7.3	-1.7	-1.6	+0.9	+1.1

erottaa-verbin taulukoita

Taulukko 31x. Muuttujan Y_SEM solukohtaiset Pearsonin jäämät. Herkkyystaso 1 = solu > 2 absoluuttisesti.

	MUUT	ERI	IRTI ₂	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-4.3	-2.7	-8.6	+2.3	+5.9	+8.7
KONKRETTIA	+3.4	-1.4	-5.8	+3.5	+3.6	-3.2
INSTITUUTIO	+8.9	+0.6	-2.8	-1.7	-4.0	-1.2
JOUKKO	-0.5	+3.4	+4.8	-2.7	-2.9	-2.3
YKSIÖ	-4.1	+0.6	+13.1	-2.2	-4.7	-4.1

Taulukko 31x2. Muuttujan Y_SEM solukohtaiset X^2 -kertymät. Herkkyystaso 3 : solu > 26.29623

	MUUT	ERI	IRTI ₂	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-9.7	-4.5	-37.6	+3.4	+16.0	+44.4
KONKRETTIA	+7.3	-1.5	-21.3	+9.5	+7.4	-7.2
INSTITUUTIO	57.3	+0.3	-5.5	-2.5	-10.1	-1.1
JOUKKO	-0.1	+8.9	+14.1	-5.5	-4.7	-3.8
YKSIÖ	-10.8	-0.3	+104.6	-3.7	-12.4	-11.9

Taulukko 33x. Muuttujan Z_SEM solukohtaiset Pearsonin jäämät. Herkkyystaso 1 : solu > 2.

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-3.3	-3.5	-8.0	+2.6	+6.7	+4.2
INSTITUUTIO	+3.4	-1.8	+11.7	-1.8	-7.3	-3.0
ELOLLINEN	+0.3	+6.8	-3.1	-1.4	-0.3	-2.1

Taulukko 33x2. Muuttujan Z_SEM solukohtaiset X^2 -kertymät. Herkkyystaso 3: solu > 18.30704

	MUUT	ERI	IRTI	HAVAITA1	HAVAITA2	LUOKITTELU
ELOTON	-3.3	-4.8	-24.1	+2.9	+12.4	+6.6
INSTITUUTIO	+6.4	-2.2	+88.3	-2.3	-25.5	-5.8
ELOLLINEN	0.0	+35.0	-7.0	-1.5	0.0	-3.1

Taulukko 34x. Muuttujan Z_MUOTO X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 15.507. Herkkyystaso 1: solu < 2 absoluuttisesti.

	MUUT		ERI		IRTI		HAVAITA1		HAVAITA2	
NA	-12.4	-5.2	-7.5	-3.5	+42.0	+9.1	+10.5	+4.1	-4.9	-3.3
NP	+9.5	+6.0	+1.0	+1.7	-14.2	-6.9	-5.7	-3.9	+0.5	+1.3
PRON_RAK	-2.0	+1.8	+8.8	+3.2	-8.8	-3.5	0.0	0.0	+7.5	+3.4

Taulukko 35x. Muuttujan RAKENNE solukohtainen X^2 -kertymä. Merkittävä solu > 23.684.

	IRTI	HAV1.	IRTI2	ABSTRAKTI	HAV2.	LUOKITTELU	KONKRETTIA	ERI
REKTIO	-4.22	-2.35	+0.009	+0.38	+0.38	-3.61	+5.55	+9.28
VAIN_Y	+95.65	+18.87	-7.07	-3.24	-17.19	-9.56	-4.85	-5.65
KONNEKTORI	-34.18	-3.25	+4.81	+0.59	+7.94	+26.61	-1.28	-3.50

Taulukko 35x2. Muuttujan RAKENNE standardoidut Pearsonin jäämät. Merkittävä solu > 2

	IRTI	HAV1.	IRTI2	ABSTRAKTI	HAV2.	LUOKITTELU	KONKRETTIA	ERI
REKTIO	+0.1	-2.2	+0.1	+1.3	+1.0	-3.1	+3.6	+4.4
VAIN_Y	-3.1	+5.1	+12.8	-2.1	-5.7	-3.7	-2.9	-2.8
KONNEKTORI	+2.7	-1.3	-8.0	+0.5	+4.0	+6.8	-1.3	-2.3

Taulukko 38x. Muuttujan X_MUOTO X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 31.41043. Herkkyystaso 1: solu < 2 absoluuttisesti.

	MUUT		ERI		IRTI		HAVAITA1		HAVAITA2		LUOKIT.	
NA	+51.0	10.2	-7.5	-3.6	+22.7	+6.9	-10.5	-4.2	-36.6	-9.2	-5.9	-3.3
NOLLA	-14.0	-4.7	-6.1	-2.8	-10.6	-4.1	+9.5	+3.5	+10.3	+4.3	+18.5	+5.1
NP	-8.2	-3.8	-0.4	+0.7	-0.4	-0.9	0.0	-0.2	+7.6	+3.9	0.0	0.0
PERS	-5.6	-2.8	+0.3	-1.8	-4.0	-2.4	+9.1	+3.2	+9.1	+3.8	+0.2	+0.5
PRON	-4.7	-2.5	+138.4	+12.6	-5.7	-2.8	-0.3	-0.6	0.0	-0.2	-3.7	-2.1

ero-nominin taulukoita

Taulukko 42x. Muuttujan Y_MUOTO X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 21.026. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI		HAV		IRTI		TEHDÄ		PUOLISO		YMMÄRTÄÄ		NUMEROT	
NA	-1.7	-2.3	+1.3	+1.4	-0.7	-1.0	-3.8	-2.4	33.1	+6.9	-4.2	-2.5	+7.1	+3.2
NP	0.0	-0.2	-1.4	-1.8	+3.1	+2.7	+1.1	+1.7	-6.6	-4.0	+2.0	+2.7	-1.5	-1.9
PRON	+3.0	+2.9	+0.5	+0.8	-4.4	-2.4	0.3	+0.6	-6.4	-2.9	-0.2	-0.6	-1.2	-1.3

Taulukko 43x. Muuttujan Y_SEM X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 12.59159. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI2		TEHDÄ		YMMÄRTÄÄ		HAV1.		IRTI		PUOLISO		NUMEROT	
ELOTON	+1.2	+3.2	+0.4	+1.2	+0.3	1.0	+1.1	2.0	-20.7	-8.8	-7.1	-5.1	+0.6	+1.4
ELOLLINEN	-3.0	-3.2	-0.9	-1.2	-0.7	-1.0	-2.8	-2.0	51.8	+8.8	17.9	+5.1	-1.4	-1.4

Taulukko 44x. Muuttujan Z_MUOTO X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 12.59159. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI		TEHDÄ		YMMÄRTÄÄ		HAV1.		IRTI		PUOLISO		NUMEROT	
ELOTON	+2.5	3.5	+7.2	4.2	+5.7	+3.7	-1.9	-2.1	-1.6	-1.9	-20.2	-6.9	-13.5	-5.7
ELOLLINEN	-3.0	-3.5	-8.6	-4.2	-6.8	-3.7	+2.3	+2.1	+1.9	+1.9	24.2	+6.9	16.3	+5.7

Taulukko 48x. Muuttujan RAKENNE X^2 -jakauma (vaalealla taustalla) ja Pearsonin jäämät (tummalla taustalla). Herkkyystaso 3: solu < 15.507. Herkkyystaso 1: solu < 2 absoluuttisesti.

	ERI		TEHDÄ_YMMÄRTÄÄ		HAVAITA		IRTI_PUOLISO		NUMEROT	
REKTIO	-23.3	-7.6	0.0	0.0	+202.6	+15.8	-8.7	-3.4	-0.1	-0.3
NOLLA	-3.0	-3.0	0.5	+0.8	+14.3	+4.5	-7.0	-3.3	+12.8	+4.2
MUUT	+6.3	+7.1	-0.1	-0.7	-44.5	-13.4	+5.7	+5.0	-3.4	-3.6